Towards the development of localised and reusable teaching and learning content and resources for languages

From one Content Database to Many Teaching and Learning Resource Outputs

Katrina A. Keogh

Ph.D.

2015

Towards the development of localised and reusable teaching and learning content and resources for languages

From one Content Database to Many Teaching and Learning Resource Outputs

Katrina A. Keogh B.Sc. (Hns.)

Ph.D.

Dublin City University

Under the supervision of

Dr. Françoise Blin

School of Applied Language and Intercultural Studies

July 2015

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy (Ph.D.) is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____

Katrina A. Keogh

Candidate ID No.: 98399195

Date: _____

Acknowledgements

To my Mum and Dad, thank you for being my teachers for as long as I can remember, for supporting me in everything I have attempted and for believing I could do it. Without your words and shows of support, I'd never have made it through.

To my husband Aylwyn, who has been there since the start, thank you for challenging me and keeping me on track, for picking up the load at home and for helping me make it to another important day in my life. To my special little people - Ewan, and twins, Alannah and Thomas - who have enriched my life more than I thought possible, and who always help me remember life's priorities. I'm looking forward to having all that time with you again and being able to stay awake while reading those books!

To my family and friends, thank you for your support, collective wisdom, open-ear and welcomed fun times. Especially to my sister Clare and friend Cara, my co-Ph.D.-ers, who know just what's involved. To Róisín, who was a constant source of optimism and encouragement and who also translated the initial draft of the linguistic resources from English into Irish – go raibh míle maith agat.

To my supervisors, Dr. Françoise Blin and Prof. Josef van Genabith, thank you for your insight, patience and support over the past number of years. Thank you for having the faith to persevere with my research and I through the many pauses life brought. A very special thank you to you Françoise, for helping to get me to the finish line and always making yourself available - while on holidays, at weekends and at all hours of the morning and night, you always responded. You supported not only the work, but the person behind it too. A truly heartfelt *merci beaucoup*.

I also owe a huge debt of gratitude to Dr. Dorothy Kenny and the staff of SALIS who provided a place for a homeless Computing Ph.D. student, and who understood that ill health, family and having children does not diminish one's ability. To the former members of the ICALL group at the National Centre for Language Technology, Cara, Joachim and Thomas, thank you for your feedback, discussion, debate and friendship. Thank you to Thomas for proof-reading the initial draft of the German linguistic content.

To my friends and former colleagues at the National Council for Curriculum and Assessment, for facilitating time to study and write, supporting the research and checking-

in! Many of you personally understood the Ph.D. process, and I found it helpful to remember that "the best thesis, is a *done* thesis!"

To the teacher, children, school and principal who were involved in this research, *vielen Dank* for the opportunity to work with you, for making space in your already busy days and for always remaining open to having German in your classrooms.

To my viva voce examiners, Dr. Phil Hubbard and Dr. Monica Ward, thank you for making time to read and consider my research, and for the interesting and thought-provoking discussion during my viva voce. To Dr. John McKenna, thank you for chairing the session and delivering that good news!

I would like to acknowledge the support of the Irish Research Council for Science, Engineering and Technology: funded by the National Development Plan. Embark Initiative Scholarship - RS/2002/441-2.

I would also like to acknowledge Ernst Klett Sprachen GmbH, who consented to my using the text in their textbook as below and provided me with an electronic textbook script in MS Word format for the purposes of this research.

Title of the book: Passwort Deutsch.
Author: Hillman Curtis
Publisher: © Ernst Klett Sprachen GmbH, Stuttgart, 2005 ff. Alle Rechte vorbehalten.

For my Dad, who always knew ‘how’ and ‘why’

Table of Contents

List of Abbreviations

AI – Artificial Intelligence

CALICO – Computer-Assisted Language Instruction Consortium

CALL – Computer-Assisted Language Learning

CL – Computational Linguistics

CSO – Central Statistics Office

DBR – Design-Based Research

DES – Department of Education and Science/Skills

EFL – English as a Foreign Language

ESOL – English for Speakers of Other Languages

ESL – English as a Second Language

FST – Finite State Transducer

HCI – Human-Computer Interaction

ICALL – Intelligent Computer-Assisted Language Learning

ICT – Information and Communications Technology

INTO – Irish National Teachers' Organisation

L1 – First Language

L2 – Second Language/Foreign Language

LS – Localised Setting

Mbps – Megabits per second: The measurement used to indicate internet speed / data transfer speed

MLPSI – Modern Languages in Primary Schools Initiative

MT – Machine Translation

NCCA – National Council for Curriculum and Assessment

NCTE – National Centre for Technology in Education

NLP – Natural Language Processing

NPADC – National Policy Advisory and Development Committee

PCR1 – Primary Curriculum Review, phase 1 (NCCA, 2005a)

PCR2 – Primary Curriculum Review, phase 2 (NCCA, 2008a)

POS – Part of Speech

PSC – Primary School Curriculum (Government of Ireland, 1999)

SLA – Second Language Acquisition

SRF – Semantic Reply Forms (c.f. Colpaert and Decoo's work)

TL – Target Language (the language being learned)

XML – eXtensible Mark-up Language

List of Tables

List of Figures

Abstract

Towards the development of localised and reusable teaching and learning content and resources for languages: From one Content Database to Many Teaching and Learning Resource Outputs

Katrina A. Keogh

Resources are a day to day requirement in all educational settings. Despite advances in the development of digital resources, paper-based resources such as textbooks and coursebooks prevail as the most commonly sought after and used resources (Gray, 2013; OECD, 2014). Three main resource challenges can be identified, (1) the need to localise global resources to their local audience and curriculum, (2) the need to develop resources in a way which allows the time, effort and expertise expended in their development to be reused and transferred to alternative formats and contexts (3) the need to ensure the separation of content from its end product to allow for content reusability.

Two discrete but overlapping areas are explored through the research. It examines (1) an alternative way to develop language learning resources which focuses on the three identified resource challenges of localisation, reusability and transferability of content. This is achieved through the four-staged development of a flexible XML database structure and a proof of concept prototype linguistic content database for teaching German to primary school children in Ireland. Classroom research was carried out in a local Irish primary school to test resource content and ensure the adequate localisation of content to the children in the classroom. Additional local resource challenges were identified and considered in the creation of the local content. As a result of this work, a new, yet untested, development model is proposed which generalises the processes and methods employed in this research to the development of localised and reusable language teaching and learning resource content databases.

This thesis also examines (2) the characteristics and makeup of the content of the prototype database revealing that the curriculum-bound database of linguistic content tailored to beginner language learners, is, in itself, a controlled language. Controlled languages are reduced and simplified forms of a language, which adhere to a set of grammatical and lexical rules in their makeup. This feature has positive implications for the database content in its ability to be automatically processed and more readily translated into other languages.

Chapter 1 Introduction

Resources, resources, resources!

Those working within the educational sector in any jurisdiction, especially those working in direct contact with primary and secondary school teachers, will all concede that one of teachers' most common requests is for teaching resources. Evidence has shown that paper-based resources are the mainstay of every educational setting (Gray, 2013; OECD, 2014) and their use as a principal resource in classrooms is increasing (OECD, 2014). While some advances have been made in introducing electronic-based or digital materials, paper-based materials such as textbooks, coursebooks and teacher manuals prevail.

However, existing resources have many challenges associated with them – the need to localise global coursebooks to make them more relevant to their local learners, their distance from the local curriculum, how they are traditionally developed and the need to maximise the time and effort employed in their development by making their content more reusable in alternative formats and contexts.

Two discrete but overlapping areas are explored through the research. It examines (1) an alternative system to develop language learning resources which focuses on the localisation and reusability of content. The local context of German language learning at primary school level in Ireland is used as the local research context, with its additional associated challenge of creating a more harmonious language learning experience for learners across the English, Irish and German languages in their sphere. It also examines (2) the characteristics and makeup of the resulting primary language resource content and how language learning content can be analysed using measures from the Computational Linguistics (CL) and Second Language Acquisition (SLA) fields.

1.1 Background

The research reported in this thesis started in the primary language classroom in Ireland. While teaching German at primary level, I found myself investing many hours developing resources to cater to the needs of the children in my classroom while teaching to the content outlined in the primary modern language curriculum - *Pilot Project on Modern Languages in the Primary School: Draft Curriculum Guidelines* (NCCA, 1999) and associated *Modern Languages in Primary Schools. Teacher Guidelines* (2001). As a linguist rather than a primary school teacher, I also needed to make a purposeful effort to integrate other relevant areas of the *Primary School*

Curriculum (PSC) (Government of Ireland, 1999) into my teaching and try to make connections between the languages already in the environment of the children in my class (English and Irish) and the language being learned (German). It became obvious that there had to be an easier way to create resources, through which I could meet the children's needs (create localised resources) as well as ensure that I could reuse resource content when it came to reinforcement and revision. I was similarly sure than I was not the only teacher who was investing such time to the same ends, so I started to investigate resourcing needs and solutions in other educational settings.

The requirement for (curriculum-based) resources

Educational settings are busy places. Teachers are required to prepare their lessons, source relevant resources and content, teach lessons, examine students, prepare and facilitate extra-curricular activities as well as ensure that they are covering the contents of their local curriculum. Searching out, editing, evaluating and developing resources is a required but time-consuming task. At the same time, resources are essential to teaching and learning. In Ireland, the reality for primary school teachers is an ongoing struggle to teach the "overloaded" Primary School Curriculum (PSC) within the allocated time of the school day (NCCA, 2005a, 2008a, 2010). Alongside the comprehensive PSC (11 subjects housed across 23 books and 2,650 pages), teachers are required to incorporate some seven sets of guidelines from the NCCA "which are intended to provide practical support to schools on specific aspects of curriculum and assessment" (NCCA, 2010, p. 6), work through school and classroom planning, report to parents, and make space for programmes designed by other organisations and agencies in support of the curriculum and initiatives (examples include Young Scientist, Seachtain na Gaeilge (Irish week to promote the use of the Irish language) and Gleo (school award for the promotion of spoken Irish)) (NCCA, 2010, p. 6). It is no surprise that stretched primary school teachers in Ireland and further afield are eager to source ready-made resources.

Paper-based resources, in particular, textbooks, have retained the dominant presence they have had through the years. References to textbooks, refer to commercially-produced consumable teaching and learning resources such as workbooks, textbooks, schemes, manuals and readers. Evidence suggests that the prevalence of textbooks in classrooms remains, with almost all OECD countries reporting to integrate them (OECD, 2014). Indeed in Ireland, almost all (95%) of the teachers surveyed in the *Growing Up in Ireland* study reporting that "pupils work individually in class using their textbook or worksheet most days or every day" (McCoy, Smyth and Banks, 2012, p. 24).

Teachers have been reported to have an *overreliance* on textbooks for planning, teaching and learning (Bolitho, 2008; DES, 2005; NCCA, 2005a, 2008a; Smiley and Masui, 2008). This has been attributed to lack of confidence in their subject knowledge (NCCA, 2008a; Prodromou and

Mishan, 2008), lack of local curriculum use when planning (Lumala and Trabelski, 2008; NCCA, 2005a, 2008a) and lack of knowledge of the curriculum of the particular subject (DES, 2005). In most educational settings, including Ireland, it is recommended that all planning for teaching should start with the local curriculum (DES, 2005).

However, relied-upon-textbooks are sometimes far removed from the curriculum they are designed to support (Harbison, 2008). Similarly, where they have been written with a local curriculum in mind, publishing processes which have market- and profit-led rubrics at play, tend to adapt content to place a greater emphasis on appearance and attractive design than end use and function (Bell and Gower, 2011; Mares, 2003; Prowse, 2011; Tomlinson, 2008). This can create much distance between the original curriculum-based materials and the finished published product.

Along with the reported overreliance on textbooks, there has been a parallel lack of resources reported in Irish schools (DES, 2005; NCCA, 2005a, 2008a). Resources were cited to be in short supply, particularly those which are "age-appropriate, modern and interesting" (NCCA, 2008a, p. 17) or indeed suitable (NCCA, 2005a). Teachers reported difficulty in accessing and assembling suitable resources and a need for resources, especially for teaching sensitive materials (NCCA, 2008a).

Benefits and challenges associated with existing resources: localisation

The benefits and challenges of textbook use as a teaching resource have been documented (Ball and Feimann-Nemser, 1986; Bell and Gower, 2011; Decoo, 2011; DES, 2005, 2005a, 2007, 2009; Harbison, 2008; NCCA, 2008a, 2010a; Jolly and Bolitho, 2011; Varley, Murphy and Veale, 2008). One of the most relevant challenges to this research is the prevalence of global, rather than local textbooks dominating in educational settings, particularly in the industry of teaching English as a foreign language. Global textbooks are those which are developed to meet the needs of large cohorts of learners internationally. As such, they do not match the local learning needs of many of their learners (Bacha, Ghosn and McBeath, 2008; Frazier and Juza, 2008) nor are they often culturally-relevant to their users (Bell and Gower, 2011; Jolly and Bolitho, 2011). Prodromou and Mishan (2008) noted that "localization will have to supplant globalization to a large degree" (p. 208) in ensuring that texts adopted are more relevant to the learners using them. Even textbooks which have been produced for a local market require a certain degree of localisation by the classroom teacher to suit the individual needs of their learners, albeit to a lesser degree than with global textbooks.

The localisation of teaching resources can also account for some local teaching and learning factors. The local Irish setting of primary school children learning German within the framework

of the *PSC* (Government of Ireland, 1999) and the *Pilot Project on Modern Languages in the Primary School: Draft Curriculum Guidelines* (NCCA, 1999) presented many local challenges to try to overcome. The Modern Languages in Primary Schools Initiative (MLPSI) was a national initiative to integrate modern languages into primary schools. It commenced in 1998 and ran to 2012, at which point it fell victim to the recession and government cutbacks. While this research has overtaken the existence of the MLPSI, the MLPSI has left a curriculum legacy as well as much research on the teaching, learning and integration of languages in Irish primary schools. These valuable resources informed this research and may become more relevant again in the near future as the NCCA consider a language education to encompass and include all of the languages of the PSC (NCCA, 2012, 2014). Children in Irish classrooms already have knowledge of English and Irish. More than 20,000 children for whom English is an additional language bring an additional linguistic wealth of approximately 167 languages into primary schools (NCCA, 2008). 17% (27,508) of primary school children across 550 primary schools also learned a modern language during the 2011/2012 school year (calculated from KEC, 2012).

In recent years, the educational sector has reapplied its focus to language teaching and learning in Ireland. Many shortcomings have been identified in Ireland's language education at all levels (e.g. Council of Europe and DES, 2008; Little, 2003; Royal Irish Academy, 2011), but in the case of primary education in particular, difficulties with literacy have been identified (DES, 2010, 2011; Perkins et al, 2010), reduced language competence in Irish has been reported (Harris, Forde, Archer, Nic Fhearaile, and O'Gorman, 2006), challenges with the implementation of the English and Irish Curriculums have been documented (DES, 2005, 2005a, 2007; NCCA, 2005a, 2008a) and a lack of a coordinated and inclusive framework for languages has been evidenced (NCCA, 2012; Ó'Duibhir and Cummins, 2012). The NCCA have recently undertaken work to develop an integrated language curriculum for early years and primary education (3-12 years) (NCCA, 2012). It is the first national effort to include all of the linguistic variables and elements present in Irish early years settings and primary schools within the same framework. Classroom resources to complement this integrated language education will be required.

The development of resources

The development of textbooks and other types of resources is carried out in different ways across different disciplines, and the process of their development is often one of the sources of challenges relating to their use. Content authoring is carried out by teacher-authors, publishing houses, the language materials writing domain and within the Computer-Assisted Language Learning (CALL) field. Each domain offers principles and/or methodologies for content development, which are examined in this thesis.

Textbook authors have been reported to rely on their own intuitions as teachers or language learners when writing textbooks, working in an ad hoc manner and calling on their creativity to develop textbook content (Harbison, 2008; Prowse, 2011; Tomlinson, 2013). The language materials writing domain has reported on principle-based authoring (Bell and Gower, 2011; Farr, Chambers and O'Riordan, 2010; Tomlinson, 2011) and methodological-based authoring (Jolly and Bolitho, 2011; Tomlinson, 2013). Templates and data-driven means, such as employing corpora (McCarten and McCarthy, 2010; Sinclair, 1986), have also been used.

Other disciplines like Computer-Assisted Language Learning (CALL) have developed robust methodologies for authoring CALL materials (content and associated software) (Colpaert, 2004; Hubbard, 1996). CALL is a multi-disciplinary field which draws on insights from Applied Linguistics, Psychology, Instructional Technology and Design, Human-Computer Interaction (HCI), Artificial Intelligence (AI) and Computational Linguistics (CL) /Natural Language Processing (NLP).

One common theme which emerges across all of these disciplines is the time consuming nature of resource and content generation (Colpaert, 1997; Delcloque, 1998; Gray, 2013, McGrath, 2002), which requires experts (Brusilovsky, Knapp & Gamper, 2006). Colpaert (2004) highlighted the importance attached to content development for CALL courseware – "most of the work in terms of time and energy was spent, not on software development, but on content development, which for most packages took one to two man-years" (p. 15). Similarly, McGrath (2002) cautioned that "the time, effort and skills required for materials development should not be underestimated" (p. 152). McGrath (2002, p. 152) cited Hutchinson and Waters, who maintained that "materials writing is best regarded as a last resort, when all other possibilities of providing materials have been exhausted" (1987, p. 125). Brusilovsky, et al (2006) reinforced the opinion that content development is a difficult and demanding process – "The development of the core content and a minimal set of educational content is a very difficult and time-consuming process, which has to be done by experts" (p. 199).

Lack of reusability in developed resources

Once language learning content has been written and completed, it can be 'hidden' in its end product and difficult to access or reuse in an alternative way. In examining the domains from which development methodologies have emerged, content could be embedded in a designed and published textbook, or form the core of a Computer-Assisted Language Learning tool. With so much time and effort invested in initial content production, it would be more beneficial for this content to be developed and made available to its end product, and potentially other products, in a reusable way. Some researchers (e.g. Decoo and Colpaert, 1997; Knapp, 2004 and Ward, 2002) have posited the concept of content databases, content cores or learning packets which isolated

the learning content from its display mechanism and could be employed in the development of language teaching and learning resources. Indeed, the topic of sustainability has come to the fore in more recent CALL publications highlighting how CALL practitioners are becoming more cognisant of the importance of transferability, reusability and future-proofing in their development work (e.g. Blin, Jalkanen, Taalas and Nic Giollamhichil, 2013; Jalkanen, 2013; Liou et al, 2013; Ward, 2013).

Analysis and characterisation of resource content

Analysing language learning content becomes important when so much emphasis is placed on using ready-made resources in the classroom. If busy practitioners are straying from curriculum documents and favouring published textbooks, then the content of these texts becomes important. Content analysis is an established function in some disciplines like SLA and CL.

The SLA domain encompasses many theories and practices for teaching and learning languages. Two rather specific areas of research in SLA involve the analysis of learner language for (1) accuracy, fluency and complexity and (2) learner errors. The former of these is usually carried out by the teacher or researcher to test whether learners are progressing in their language learning, but have also been put to use in related disciplines such as learner modelling in CALL (Heift and Schulze, 2007) and assessment of language impairment and delay (Malvern, Richards, Chipere and Durán, 2004). Accuracy examines how correct the language produced by learners is, fluency examines how native-like learners' language is and flows, while complexity examines the range of words (lexical) and structures (grammatical/syntactic) which learners use and can call on from their language repertoire (Wolfe-Quintero, Inagaki and Kim, 1998). These measures are calculated by using mathematical ratios.

The second of these specific areas from SLA, informs ways to describe and classify *learner errors* as another means to understand learners' levels of competence and performance in a target language (Granger, 2003; Thouësny, 2011). Information gleaned from learner error analysis can in turn feed back into the teaching and learning process (e.g. through feedback loops as described in Schulze 2003).

CL, sometimes called Natural Language Processing (NLP) is a domain which develops tools and systems to automatically process and analyse natural languages. Analysis within CL may be required to break words down to their stems, attach their parts of speech (POS), examine the syntax of a sentence or join up multiple segments of words to form new words (e.g. plurals). These and further processes can be combined to form tools for different natural languages such as POS taggers, parsers, speech processing tools, machine translation systems and question-answering systems. Controlled languages are subsets of language, which are bounded by

particular grammatical and lexical rules. They are used in the Machine Translation (MT) industry (among others) to limit translation input to the degree that it can all be accounted for in automatic translation.

Virtual resources and ICT infrastructure

While paper-based resources dominate, scope should also be provided for virtual or electronic resources. Teachers do presently integrate ICT, but the majority of those who do, do not use ICT to its full potential (DES, 2005, 2005a, 2007, 2008, 2009; NCCA, 2005a, 2008a) and few teachers and schools plan for the integration of ICT (DES, 2009). Time constraints, lack of classroom management strategies, insufficient ICT infrastructure, lack of teacher confidence and lack or unavailability of suitable resources were cited as reasons for low uptake and use of ICT by teachers (NCCA, 2005a). Children were reported to enjoy ICT-based activities (NCCA, 2005a), but their levels of ICT ability were reported to be low (DES, 2008).

It is evident that schools and teachers require a broad spectrum of resources in order to implement the PSC. There is a need for these resources to support curriculum implementation, teachers' lack of subject knowledge in some cases and children's learning. Resources require much teacher time to source, evaluate, develop and adapt. It is necessary to ensure these resources are catered to and differentiated for the individual needs of children in primary schools.

With ready-made resources and textbooks prevailing and providing assistance for stretched teachers, and with a plethora of development strategies at play, a new and comprehensive method was required to develop a curriculum-based and relevant local resource in a systematic and robust fashion while ensuring content could be localised to its target learners and reused in other settings.

1.2 Research questions

This research sought to create a specification of the structure of a linguistic content database, which would enable the creation of language teaching and learning resources that are reusable and transferrable while addressing local teaching needs (localised). In doing so, it developed, populated and tested a prototype/proof of concept database for the German primary school context with a view to determining the following:

1. How can language learning content and resources be localised to particular learning contexts?
2. How can resource content be reused and transferred to other language learning environments?

3. How can a generic linguistic content database be structured?
4. Does linguistic content bounded by a curriculum and localised settings exhibit features of a controlled language?

1.3 Stages of work

Two focuses

The research questions outlined in Section 1.2 are explored through two distinct focuses of work, which are not in themselves, mutually exclusive. The first proposes an alternative system for developing language learning content and resources, while focusing on the localisation and reusability of content. This is achieved through the development of a flexible XML database structure and a proof of concept prototype linguistic content database for teaching German to primary school children in Ireland. Classroom research was carried out in an Irish primary school to map the requirements of a language learning environment to the specification of the linguistic content database. In addition, the content for the prototype database was tested to ensure localised language learning needs and challenges could be addressed through the content.

The second focus is on examining the characteristics and makeup of the content of the prototype database to ascertain how the database-bounded linguistic content of a beginner's language learning curriculum, differs from more traditional learning resources like textbooks and authentic language. Similarly, it is examined to explore whether it exhibits any features of controlled languages, which would have far reaching implications for its ability to be processed and translated automatically.

Four phases of work

These two focuses were progressed through four phases of work. The first phase required that preparatory work was completed before entry into the classroom. The relevant localised context for the Irish primary school modern language learning setting was broken down across five localised settings. *Localised settings* were used to account for the features which make up the whole of any one context and which are to be taken into account in the development of localised learning content and resources. The five identified for the modern language primary classroom were:

1: The Primary School Curriculum (Government of Ireland, 1999)

2: Linkage and integration topics across the PSC and the modern language curriculum documents

3: The two modern language curriculum documents - *Pilot Project on Modern Languages in the Primary School: Draft Curriculum Guidelines* (NCCA, 1999) and associated *Modern Languages in Primary Schools. Teacher Guidelines* (2001).

4: Language awareness

5: Typical learner errors.

The German modern language classroom of an Irish primary school was used as the research site over a period of two and a half years to iteratively present, test and gather data on the five localised settings. The results of this work presented (1) a specification for a linguistic content database, as well as (2) content to populate the specification for a localised instantiation of the database.

The second phase of work employed methods of analysing text and content from the SLA and CL domains. While this phase did not contribute to the development of the database specification or prototype, it did examine the content which would be used to make up the prototype. The aim of this phase of work was to ascertain whether the content devised for the prototype linguistic content database differed from other language learning resources like a textbook or authentic language found in a newspaper. A further aim was to characterise the language contained in the database.

The third phase involved the construction and population of the prototype linguistic content database. The final fourth phase explored proof of concept resource outputs such as language learning handouts and worksheets which can be automatically generated by the teacher/practitioner from the linguistic content database.

1.4 Scope of the thesis

The research carried out across the four phases of work crossed many disciplines. The overall development methodology employed was informed by Hubbard (1996) and Colpaert's (2004) development methodologies. Their work inspired me to undertake the development of the localised linguistic content database, as I could see how a solution could start to be realised while taking all of the local settings of the primary school classroom into account. A three-pronged methodological approach was taken to inform the four phases of work. Phase I had its basis in Design-Based Research (DBR). Design-Based Research is sometimes called development research, and is a mixed-methods approach to research in a target space. DBR was defined as "an emerging paradigm for the study of learning in context through the systematic design and study of instructional strategies and tools" (DBRC, 2003, p. 5). Phase II represented a Computational Linguistics (CL)/Natural Language Processing (NLP) study which drew on

methods and tools from these domains. Phases III and IV represented the development work which drew on programming and software engineering methods as well as Hubbard's (1996) and Colpaert's (2004) CALL methodologies as described.

Within Phase I, the teaching focus and the focus on (second) language learning in the classroom was informed by the local context and the curriculum which bounded the learning of the children participating in the research. Research methods employed in the classroom represented a mixture of qualitative and quantitative means, reflecting the DBR ethos. Preparation and data gathering surrounding error analysis was informed by the work of Granger (2003) and Mackey, Gass and McDonagh (2000). Error post-processing and analysis were informed by the work of Granger, Kraif, Ponton, Antoniadis and Zampa (2007), L'Haire, (2007) and Thouësny (2011).

Phase II drew on work from the SLA and CL domains, including Wolfe-Quintero, et al (1998) who provided a meta-analysis of SLA matrices for calculating accuracy, complexity and sophistication in learner's production. Phases III and IV drew on the work of Colpaert (2004), Hubbard (1996), Knapp (2004) and Ward (2002).

The research in this thesis also builds on work from Decoo (2011), Maley (2011) and Prabhu (1988) who have all provided accounts of extracting or abstracting language learning content from its display mechanism and returning to words and phrases as the basis of resource content generation – what I have termed *raw linguistic content*. This work aims to quantify and delineate the content which would be required to reach a certain level of proficiency in a modern language, bounded by a local curriculum. It sits alongside the variable factors of learning that the teacher, learner or other actor/factor bring to the setting.

This approach is more in keeping with a computational linguist's view of language, in its effort to quantify what is usually unquantifiable, or produce a language learning code which is to be cracked by language learners in order to reach that level of proficiency. It is a viewpoint which is at odds with how language teachers and theorists view the language being learned by the individuals in their classrooms. It is also at odds with the task-based language learning classrooms of today's educational settings where learning outcomes are based on skills rather than the acquisition of particular vocabulary or phrases. However, it is in line with how resources currently operate in educational settings, and with how resources themselves are finite in their content. They delineate what is to be or can be learned for the target audience. In resource generation, it is important that it is the teacher/practitioner who brings the didactics to the learning, rather than the materials which should instead act as a supporting mechanism. This approach aims to empower the teacher to act as the didactic professional and place them in the "materials-power" position, rather than the resources, by providing a source of raw linguistic

content from which teaching and learning resources can be generated. With a background as a linguist, computational linguist and language teacher, it is hoped that the results of this research will sit comfortably with practitioners from each of these disciplines.

It is proposed that as a result of this research, the processes and methods outlined could be generalised to other contexts in the form of a new development model. This development model, called the Tailor Development Model, aims to support other researchers in similarly localising content to their specific contexts and developing localised linguistic content databases.

1.5 Thesis structure

Following a definition and categorisation of educational resources, Chapter 2 discusses those educational resources which are most relevant to this research. It focuses in on the most dominant type of resource in educational settings – the textbook. Textbook use, control and production methods are explored along with the associated benefits and challenges of their use. The second half of this chapter examines other types of resources such as structural and physical ICT resources.

Chapter 3 describes existing language learning resource development methodologies from the language materials development domain and CALL. It also explores how content can be processed and analysed to reveal its characteristics and features. This chapter concludes by mapping out the global problem research space.

Chapter 4 introduces the local context of the research, and the additional local challenges which form part of the research space. It describes the background to the local settings for the primary German language classroom in Ireland. It outlines the research design and methods employed through four phases of work.

Chapter 5 explores Phase I of the research work which was largely conducted in the primary school classroom. Results from data gathering in the classroom are presented and the specification for the linguistic content database is outlined.

Chapter 6 examines how the content of the linguistic resources were processed, analysed and characterised in Phase II of the work. It defines controlled languages according to rules and describes analysis matrices from SLA. Three datasets are introduced, each of which is analysed according to controlled language features and SLA matrices. Analysis results are presented.

Chapter 7 traces the processes of Phase III which involved converting the raw linguistic content to a linguistic content database. This process required moving the data from Excel to XML and adding annotations referencing the data to student errors, the Primary School Curriculum, multimedia resources and other languages being learned by the children in the primary school classroom. Chapter 7 also describes the outcome of Phase IV which provides proof of concept applications which draw on the linguistic content database. These applications exhibit how the didactic database can be reusable and accessible. Further examples of application are posited.

Chapter 8 returns to the research questions and provides conclusions, limitations and recommendations for further research. The Tailor Development Model is also outlined at this final juncture. As a generalisation of the research design and an outcome of this research, the resulting model can be further tested and validated in other research contexts.

Chapter 2 Resources in teaching and learning

This chapter provides a thorough review of resources in educational settings, in particular, consumable paper-based resources (the textbook) and digital resources. The textbook dominates the discussion in this chapter as it is one of the most widely discussed and accessible ways to gain an insight into the status quo of resourcing in educational settings. Resources are the mainstay of any educational setting. They are used to support the day to day teaching and learning which takes place in these settings. There is a sensitive balance between the availability of resources, their relevance to any one target learning environment (including their need for localisation) and the time a practitioner has available to search out, evaluate, edit or develop relevant materials for their learners (or indeed, for learners to undertake these processes for themselves). Textbooks present their own merits and challenges, and while they are a staple resource in educational settings, their 'over-integration' can be a cause for concern in some instances.

Digital resources similarly present benefits and challenges to educational settings. Their most relevant feature in the context of this research is how the localisation of digital resources aligns to the localisation of software from the translation industry. The latter provides insights which inform this research.

This chapter first defines and categorises teaching and learning resources, pinpointing those which are most relevant to this research. The following sections then narrow the resource focus to consumable paper-based resources (Section 2.2) and digital resources (Section 2.3), and their standing across themes such as status, control and supply, standardisation and over-reliance. Section 2.4 briefly examines the place of structural and physical resources relating to ICT, as these resources can provide the framework for the integration of digital/electronic consumable resources. Section 2.5 provides an in-depth discussion surrounding the localisation of consumable resources and the actors and factors which can impact and initiate the need for this process.

2.1 Defining resources

When we refer to resources, we refer to the artefacts used to support teaching and learning in any arena. Within this research, references to resources refer to those materials which

are used directly in the teaching and learning of languages or the surrounding infrastructure which is equally important (e.g. internet, computers). The words *materials*, *resources* and *equipment* have all been used in the literature to refer to the same entity.

Levy and Kennedy (2010) used the word "materials" within their 2010 paper to refer to the learning resources that are created within Computer-Assisted Language Learning (CALL). They specified "'materials' will be used to encompass the different kinds of materials, software, courseware, programs, exercises, and learning environments that are created in CALL. Importantly, this label is used to include both content and process materials" (Levy and Kennedy, 2010, p. 529). Content data is further defined as "sources of data and information" and process material as "guidelines or frameworks for the learners' use of communicative knowledge and abilities" (Levy and Kennedy, 2010, p. 529 citing Breen, Candlin and Waters, 1979).

Tomlinson (2012) similarly defined *materials* as anything that can be used to facilitate the learning of a language, including "coursebooks, videos, graded readers, flash cards, games, websites and mobile phone interactions" (Tomlinson, 2012, p. 143). McGrath (2002) specifically defined "*text* materials" as those which have "been either specifically designed for language learning and teaching (e.g. textbooks, worksheets, computer software); authentic materials (e.g. off-air recordings, newspaper articles) … teacher-written materials; and learner-generated materials" (p. 7).

A meta-view of the definition of resources shows how resources are one of the four components required to make up a "pedagogical core" (OECD, 2013, p. 23). The term pedagogical core was composed to try to overcome the "institutional" connotations attached to 'school' and 'classroom' as educational settings (as learning can take place in other domains also) (OECD, 2013, p. 23). The more traditional educational triangle included learners, teachers and content. The OECD's Educational Core added a fourth tangible element – resources – to account for the physical requirements in any learning environment or core. Resources in this instance were defined as elements which "can be directly exploited in learning – i.e. physical resources (buildings, facilities, and infrastructure) and learning materials" (OECD, 2013, p. 23). Figure 2.1 below illustrates the OECD concept for inclusion of resources. The elements "pedagogy" and "institution" are those which link the items comprising the core (OECD, 2013, p. 24).

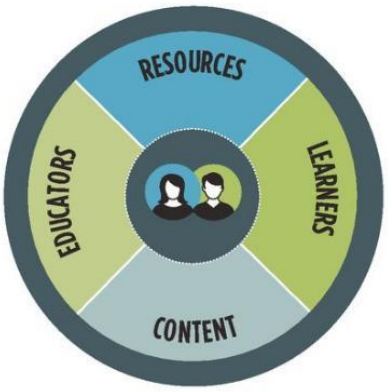


Figure 2.1 The Elements of the Pedagogical Core (OECD, 2013, p. 23)

In focusing in on the research niche of Irish educational settings, in particular the primary school setting, resources refer to the teaching and learning materials required to implement the Primary School Curriculum (PSC) and support children's learning.

Table 2.1 provides one way to categorise the plethora of resources used globally and in Irish educational settings, and those referred to in the associated literature. It represents those resources required on an annual and daily basis in all educational settings, as well as those which are employed to support them.

Resource category		**Consumable resources** (subject-specific in most instances)		**Reusable resources** (subject-specific in some instances)
		Examples		
	Paper-based	Printouts, photocopied pages, handouts, worksheets, workbooks	Posters, teacher manuals, books, flashcards, graded-readers, schemes, dictionaries, textbooks	
	Virtual/digital	Electronic exercises, SMS	PowerPoint presentations	Websites, corpora (collections of electronic texts)
	Concrete objects		Magnets, CDs, DVDs, scissors, games	
Structural and physical resources		Computers, printers, internet access, school network(s), classroom		
System resources		Human resources (including teacher time) Financial resources		

Table 2.1 Educational resources by category

A continuum of resource-types can be described, with the two end points comprising consumable and reusable resources. The grey shading between 'consumable' and 'reusable' indicate the position of the resource examples (e.g. printouts, websites) on the

continuum. Consumable resources are those which are *consumed* in the classroom and created or sourced by teachers daily/weekly/monthly/annually to support the needs and learning of their individual classes. Some of the paper-based consumable resources are reusable from one year to the next where they do not become outdated, damaged or obsolete in the case of a change to the curriculum (e.g. textbooks) and others are exhausted within their first use (e.g. workbooks and worksheets). The virtual/digital consumable resources can be reused from one year to the next, as they do not become exhausted through use. They are however, also prone to becoming outdated as textbooks are.

Reusable resources are those which will take longer to outdate and usually only need to be replaced in instances where they have become worn or damaged. As with consumable resources, their integration into teaching and learning tends to be on a subject-specific basis, with magnets, for example, being relevant to a Science Curriculum.

Structural and physical resources are required to support the use and integration of some of the consumable and reusable resources. For instance, computers will be required to display electronic exercises or for using software CD-ROMs or DVDs. These resources provide the infrastructure for teaching and learning. System resources are those which are required but are often less visible as tangible assets in the educational sphere or in learners' day to day encounters. Both structural/physical resources and system resources are generic, not being specific to any one subject.

Further categorisations have been offered to help to define resources further. Tomlinson (2012) offered a categorisation of paper-based resources according to their function:

> "Materials can be informative (informing the learner about the target language), instructional (guiding the learner in practising the language), experiential (providing the learner with experience of the language in use), eliciting (encouraging the learner to use the language) and exploratory (helping the learner to make discoveries about the language)" (p. 143).

Gray (2013) added to Tomlinson's categorisation, stating that materials can also be "cultural artefacts" (p. 3) which tell the tale of any one set of cultural norms, be they stereotypical or not.

This research focuses on the resources within the first category–consumable resources. Structural and physical resources are equally relevant, as they are required to support the

use of consumable resources in some instances. However, their use and availability is determined by the educational institution and system. Having defined and categorised educational resources, the discussion now turns to examining the staple consumable paper-based resource of educational settings – the textbook.

2.2 Consumable resources: paper-based resources

"*Say loudly to schools and teachers … do not be slaves to textbooks!"*
(Teacher quote on how to alleviate curriculum overload, NCCA, 2010, p. 52)

The focus of discussion in this section remains with resources, but narrows to the category of one particular type of consumable paper-based resource – the textbook, and consumable virtual resources. Much empirical data about textbook use in educational settings globally and in schools in Ireland has been recorded and documented, where data on the use of other equally specific types of resources like virtual online exercises is less prevalent. A review of this research is provided below, which outlines the *conflicting messages* which emerge in relation the place of textbooks as friend or foe in the education system. The discussion provides a firm basis for the rationale for the requirement for resources in educational settings and the associated challenges which need to be addressed in any new resources being created.

When we refer to textbooks, we refer to the commercially-produced teaching and learning resources which are the basic requirement of many schools. They can take the form of workbooks, schemes, manuals and readers.

It has been difficult to compile data from other countries which all relates to textbook use at the same level of education. As a result, some of the information below relates to primary and secondary level education, while other information relates to textbooks for teaching English as a Foreign Language (EFL) or English to Speakers of Other Languages (ESOL). However, the purpose of examining textbook use at any of the educational levels is to highlight their prevalence in educational settings as well as issues which arise with the use of globally produced texts. In each case below, the level of education (e.g. primary, secondary etc.) that the discussion pertains to will be indicated.

2.2.1 Textbook requirement and status

All of the literature reviewed is unanimous on the continued use and integration of textbooks in educational settings. The OECD (2014) provided evidence of the prevalence of textbooks from data gathered across three collection points (2003, 2007 and 2011). Findings revealed that the levels of textbook penetration were increasing in some jurisdictions (e.g. Italy and Slovenia) for subjects like Science, Mathematics and reading assistance and decreasing in others (e.g. England) for these same subjects. However, in the majority of cases analysed, the OECD mean showed an increase in the use of textbooks as a primary resource (OECD, 2014).

Bolitho reflected on the core place of textbooks in Central and Eastern Europe where there is much variety across countries in the control that the government exerts on textbook selection. He concluded that "all the signs are that textbooks will be needed in English classes for some time to come, whether as stand-alone tools or as part of a more sophisticated multi-media package" (p. 221). Bacha, Ghosn and McBeath (2008) similarly reported on the textbook playing "a central role in the teaching and learning process" (p. 284) in primary schools in Middle Eastern countries. Indeed, Gray (2013) confirmed the *enduring centrality* [of textbooks] *in classrooms around the world* (p. 2). The extensive educational publishing sector lays testament to this fact, as does the extent of the literature available on materials development, in particular, for language learning, which is of most relevance to this research.

Ireland is no different from other jurisdictions, with the majority (between 50% and 87%) of teachers reporting to use textbooks "frequently" across Science, SPHE and Irish (NCCA, 2008a, p. 120). More recent evidence reinforces the dominance of textbooks and reading schemes in classrooms, above and beyond any other teaching/learning resource (Gilleece, Shiel, Clerkin and Millar, 2012). Gilleece et al (2012) reported that 98%-100% of teachers reported using textbooks for Mathematics and 92%-98% reported using reading schemes for English at least weekly in Irish-medium schools (p. 107).

In some cases, the theory behind the regard for textbooks does not always match what happens in practice. In the USA for instance, teacher training colleges were reported to highlight textbook deficiencies for trainee teachers whereby they were encouraged to use their own knowledge and professionalism to plan and source content for their lessons (Ball and Feiman-Nemser, 1986[1]). Teachers were generally taught that "they are not really doing

[1] Although dated, this remains the conflicting scenario in many educational settings (e.g. Frazier and Juza, 2008 provide similar evidence).

their job if they 'simply' take the textbook and teach straight from it day after day" (Frazier and Juza, 2008, p. 187). The reality in teaching practice was that trainee teachers needed to conform to the local culture of textbook- and manual-based teaching (Ball and Feiman-Nemser, 1986).

In some jurisdictions, where the cost of textbooks can be an inhibiting factor in providing choice and supplementary materials to students, one textbook can become the core source for all teaching and learning which takes place. In Kenya, for example, one English language textbook–the *Integrated English* coursebook–has become the staple, so much so, that it is seen as "some kind of holy book" (p. 235) from which teachers and students are reluctant to stray. Similarly, Smiley and Masui (2008) reported on the status of textbooks in Japan. They described how "the textbook holds the paramount place in the language classroom. It is the main focus of language education and is expected to provide a path that enables the learners to become competent English users" (p. 256).

Other systems seem to diligently and "rigidly" follow a textbook, where teachers have been reported to ensure that they "cover all of the material" and students "adhere very closely to the text and almost memorize it" (Bacha, Ghosn and McBeath, 2008, p. 291). Some systems have shown how parental pressure on teachers to complete textbooks which are in place can result in teachers also feeling under pressure to cover every page of a textbook. Teachers in Ireland cited parental expectations of their children's textbooks being completed within the school year as adding pressure to and driving textbook use (NCCA, 2005a). Parental pressure was also a finding for Harbison (2008) who reported that teachers felt parents defined good and bad teachers by how much of the textbook they completed during the school year. Teachers ascertained that they could make subjects more interesting by "not depending on the textbook as much" (NCCA, 2005a, p. 178).

This evidence suggests that the requirement for paper-based resources like textbooks remains, even in a time where virtual and social resources dominate outside of educational settings. The presence of textbooks can have a negative connotation where their content dictates what is taught and how teaching should progress – "textbooks were never intended to be a straightjacket for a teaching programme in which the teacher makes no decisions to supplement, to animate or to delete" (Bell and Gower, 2011, p. 138). More flexibility in resource selection, generation and use would eliminate these challenges. Conversely, varying levels of control are exerted over resource selection and use in some educational jurisdictions.

2.2.2 Control and supply of textbooks

Textbooks are a valuable teaching and learning resource if used appropriately in an educational setting. Ministerial control of textbook design and/or use varies across countries. In some countries, such as Japan, Singapore and Hungary, the state takes charge of producing textbooks (Bolitho, 2008; NFER, 1996; INCA, 2013). Other countries such as Germany, France, some USA states and Switzerland allow textbooks to be produced commercially but publishers must follow state or local government guidelines relating to content, format, quality and/or cost (NFER, 1996). Other countries (e.g. half of the states in the USA, Kenya) supply a prescribed list of textbooks, where the list is compiled by local government or the state (Collins, 2012; Lumala and Trabelski, 2008). Approved textbooks are usually those which adhere to the curriculum. Countries such as England, Sweden and the Netherlands allow teachers to freely choose which textbooks to use in their classrooms (NFER, 1996).

In the past, the Irish education system exerted control over textbook production. This practice has since been discontinued and today Ireland's textbooks are produced commercially. The National Council for Curriculum and Assessment (NCCA) does provide some guidelines for publishers on how to produce curriculum-based textbooks for teaching and learning any subject within the Primary School Curriculum. The NCCA also makes itself available to consult with any publishers on an on-going basis (NCCA, 2001a). One example provided by the NCCA for English (NCCA, 2001a) outlined the strands of the English curriculum and emphasised the importance of "congruence between the textbooks and the principles of the revised curriculum" (NCCA, 2001a, p. 1). It provided details of the methodologies to be employed and how the curriculum is based on "up-to-date research and best theory and practice" (NCCA, 2001a, p. 2). The remainder of the specification was divided between general guidelines and more specific details of the volume of resources required for each level in the Primary School Curriculum. The general guidelines comprised:

- General principles such as making the content interesting to children, basing content on current thinking as to how children learn language and ensuring content is instructive so that children will learn from it.
- Methodologies which are based on those outlined in the curriculum teacher guidelines.
- General considerations including gender equity, multiculturalism, heritage, the role of parents, the European context and the use of ICT.

The specific details of the volume of resources which should be produced were based on principles for resources (NCCA, 2001a). There was however, a caveat, that the exact numbers of books, types of books and number of pages in each book specified for each level in the Primary School Curriculum were flexible if the resource principles supplied were adhered to. More recently, research carried out through the second phase of Primary Curriculum Review (PCR2) (NCCA, 2008a) highlighted teachers' difficulty in sourcing relevant resources and their overreliance on same. In response to these findings, one of the recommendations on the use of teaching and learning resources was that "control options for textbooks for schools should be explored" (NCCA, 2008a, p. 218).

Control over textbook production points to a more cohesive and curriculum-aligned experience across education systems, but only where sanctioned resources are aligned to the curriculum and relevant to their localised setting. Furthermore, exerting control over resources similarly reduces the teacher's/practitioner's control over teaching and learning in their classrooms and the ability to localise resources to their individual learner's needs.

2.2.3 Textbook overreliance–why and why not?

The overreliance teachers have been reported to have on textbooks is present for a number of reasons which will be discussed below. This overreliance is a basis for concern in the eyes of some textbook consumers, and a comforting base for others who may lack subject knowledge in one area or are too stretched on time to plan their teaching from the curriculum.

Lack of time

Teachers have reported to be overloaded and overburdened in their day to day activities (NCCA, 2005a, 2008a, 2010). As discussed in the Irish context in Chapter 1, this has been attributed to their busy daily schedules, the paper-heavy curriculum and supporting documents, as well as the many educational programmes in place. Teachers are short on time, and have been reported to rely too heavily on textbooks and supporting resources to compensate for this shortage of an important system resource (teacher time) (NCCA, 2010). There is a sense that teachers should "spend their valuable time more on facilitating learning than materials production" (Bell and Gower, 2011, p. 135).

Lack of subject knowledge

Teachers in primary schools are required to teach many subjects comprising their curriculum. Teachers have reported to sometimes have an overreliance on textbooks, as an aid in supporting their lack of subject knowledge in any one of the many subjects present

in their curriculum (NCCA, 2008a).Textbooks also offer a support where only one subject is required of the teacher, as seen in Greece where teachers of English are Greek English L2 speakers (Prodromou and Mishan, 2008), and similarly in Japan, where teachers rely on textbooks to support the course of their teaching (Smiley and Masui, 2008).

Used as a source of standardisation in learning

Commonly used textbooks can sometimes be put to use in educational settings to standardise what is being taught from one class and year to the next, especially where such information can be lacking from curriculum documents. In using the textbook content as a basis for the course of teaching and learning, the curriculum documents are overlooked. Basal graded readers are popular in the USA and hand in hand with teacher manuals, help children to progress their reading across the class levels. Teachers in Ireland have also been reported to use textbooks as a level of standardising what is being taught across levels, also when it comes to reading (NCCA, 2005a). Teachers reported feeling secure in knowing that by using certain textbooks and readers, all classes were aligned to the same reading level (NCCA, 2005a). Similarly in the case of Gaeilge/Irish, textbooks were used to ensure progression in the language being learned across the levels, without much reference being made to the curriculum documents (DES, 2007). These issues could be exacerbated as school book rental schemes are increasingly being put in place in schools in Ireland (Barnardos, 2012; DES, 2012). If a particular series of textbooks and workbooks become the norm at different class levels, it becomes easier for teachers to use their content as a basis of standardisation rather than the curriculum.

However, Decoo (2011) has provided evidence that textbooks are rarely systemised in their approach to presenting and progressing content across class levels. After a thorough examination of textbooks, he found that they were lacking in "grading and sequencing" content from year to year (Decoo, 2011, p. xvii). He reinforces the position that the dynamics of changing teachers across school years and "haphazard focus-on-form [in] … the timing of [covering] grammatical topics" (p. 274) militates against curriculums or textbooks providing adequate coverage for coordinating learning across school years. It seems that some textbooks do not merit being used as a source of standardising teaching across levels.

Replaces curriculum-based planning and teaching.

Research from Ireland has shown that textbooks are not only used as a learning resource, teachers have reported to also use textbooks as a basis for their planning and the course of learning for a teaching year (DES, 2007; Gilleece et al, 2012; NCCA, 2005a, 2008a). In

extreme circumstances, textbooks have been described as informing whole-school planning (DES, 2005). Teachers cited textbooks as serving the function of guiding curriculum implementation. They described their satisfaction in knowing that by following textbooks they were addressing curriculum strands and strand units (NCCA, 2005a) and providing "a sound structure for their teaching" (Harbison, 2008, p. 172). Teachers have been reported to rely on textbooks in place of curriculum documents, as the sheer scale of the "Physical Face of the Primary School Curriculum" can make it difficult to navigate (NCCA, 2010a, p. 9). While some efforts have been made to remedy this problem through re-presented and simplified overviews of curriculum documents (NCCA, 2005a, 2010), and supporting guidelines for planning (e.g. from the DES, NCCA, PDST and the National Induction Programme for Teachers (Primary), 2010/2011), teachers still revert to the textbook (NCCA, 2010).

This situation is not unique to Ireland. Tomlinson (2008) also reported on an international survey on the popularity of particular coursebooks for EFL teaching. Teachers responded by saying they used one particular book as it "meant they didn't have to spend time preparing their lessons but that they felt sorry for their students because it was so boring" (Tomlinson, 2008, p. 4). Similarly, Bolitho (2008) discussed the use of textbooks in Central and Eastern Europe and outlined how the "textbook, in the hands of the teacher, was effectively the syllabus" (p. 213).

Conversely, other teachers were reported to welcome the opportunity to get away from the English textbook and reader that the revised English Curriculum in Ireland facilitated (NCCA, 2005a). They saw textbooks as "impediments to cross-curricular learning" (NCCA, 2005a, p. 236). Teachers have reported that sometimes, curriculum-based texts can be "unattractive and outdated" resulting in a "poor stimulus for learning" (NCCA, 2005a, p. 236) which did not always present authentic learning materials or materials which children today can relate to (NCCA, 2005a). One school in Ireland described going as far as abandoning textbooks, noting "the perceived overload that textbooks can generate in contrast to the potential for developing real learning using a range of good resources" (NCCA, 2008a, p. 120). Thornbury (2013) similarly reported on the "Dogme ELT" (p, 217) movement which espouses a textbook-light or "materials-light" (p. 218) approach to teaching and learning in textbook dominant classrooms. He suggested alternatives to the standard textbook-driven lessons which could be achieved by customising the textbook, by using glocal resources, or by replacing textbooks with content-based approaches to teaching and learning (Thornbury, 2013). He described how the latter of these can be

achieved through a number of means including the use of L1 materials for the TL or sourcing materials online (Thornbury, 2013).

In Irish primary schools, it is recommended "the starting point for classroom planning should be the specific objectives of the curriculum, rather than textbooks" (DES, 2005, p. 52). Furthermore, some curriculum documents specifically explicate that lessons "should not be work-card or textbook-based" (Government of Ireland, Science Teacher Guidelines, p. 27). Without any assurance that textbooks are based on the Primary School Curriculum (PSC), teachers should exercise caution in their use for planning or their use in teaching without prior planning.

One further downside of textbook-based classrooms is their reduction in the range of teaching methodologies in use. Evidence from Ireland described the detrimental effect textbook over-reliance had on the learning experiences for children. Sources cited how textbook-based reading, writing, teacher questioning and whole-class teaching activities dominated (DES, 2005, 2007; Varley, Murphy and Veale, 2008) at a cost to other teaching activities and methodologies like

- problem-solving, active learning and reasoning (Mathematics) (DES, 2005)
- "personal reading, silent reading and reading for enjoyment" (Gaeilge/Irish) (DES, 2007, p. 66).
- hands-on, activity- and investigation-based activities (such as *work scientifically* and *design and make*) (Science) (Varley, Murphy and Veale, 2008).

In addition, over-reliance on textbooks led to a limited access to a variety of genres and a more modest repertoire of known stories and poems (Gaeilge/Irish) (DES, 2007).

These textbook experiences were "evidence of negatively construed science experiences dominated by textbooks and workbooks" (Varley, Murphy and Veale, 2008, p. 152). Textbooks can be lacking in exercises, both type and quantity, as well as promoting whole-class teaching. Whole-class teaching, while useful in some teaching instances, warrants against taking the needs of each individual child into account (DES, 2005a). This, in turn, while having a negative implication in any setting, can have a more pronounced implication in settings which have been designated as disadvantaged, have a large proportion of children for whom English is an additional language or where there are any/many children who require individual learning support.

The further result of a reduced use of methodologies and catering to children's individual needs, was that children were not progressing in the learning or mastering skills which were

outlined in curriculum documents (DES, 2005, 2007). It was not just a reduced methodological experience for children which was reported, but also a shortfall in learning and achieving all that they should, as set out in the curriculum (DES, 2005, 2007) – a reduced learning experience.

The cautionary report from the DES (2005) on the over-use of textbooks is provided below:

> "Textbooks exert a dominant influence on teaching and learning in a significant number of classrooms. In these class settings the teaching tended to be didactic, and undemanding and repetitive learning tasks were provided for the pupils. There was little emphasis on the development of higher-order thinking skills, on nurturing pupils' creativity, or on encouraging pupils to respond emotionally and imaginatively. Teaching methodologies were restricted, and the essential emphases of the curriculum were not accorded due prominence. The quality of pupils' learning was found to have significant scope for development in these instances, and the pupils were not sufficiently interested or engaged in their learning."(p. 49).

McGrath's (2002) summary of the reasons teachers require coursebooks echoes the themes emerging in this discussion:

- "It provides a structure for teaching [planning]
- It saves time. To prepare materials from scratch for every lesson would be impossible [lack of time]
- It offers linguistic, cultural and methodological support [replacement for local curriculum]
- It is easy to keep track of what you have done and to tell others where you have reached [standardisation/reporting]… " (p. 11).

Throughout the examples provided above, across many educational sectors and environments, paper-based consumable resources remain dominant. In countries where there are fewer financial resources for teachers and learners to source additional texts, one or two texts become the prescribed and dominant core text for all teaching and learning. In some instances, these core texts have been localised, while in others, they are culturally unspecific global coursebooks. While unavoidable, the presence of one core book can have many pros and cons attached – teachers become familiar with one text but classes can become "unexciting and book-centred" (Lumala and Trabelski, 2008, p. 236), teachers can use the text to plan their teaching but classes can become uninteresting to learners when not supplemented with additional materials and teachers can ensure continuity across learning when one coursebook is used but a coursebook should not replace a curriculum in

accounting for progression in learning. Sometimes a local requirement or preference can filter up to the global coursebook producers who adapt a staple publication to allow for this localism (as exampled through the Greek system's preference for grammar-based texts).

As the section headings and discussion have indicated, the emergent themes are (1) textbooks are a core staple across global educational settings, (2) there has been a shift from global textbooks to a need for more localised publications and (3) teachers have an overreliance on textbooks which leads to a reduced and limited learning experience for learners. A summary of relevant findings is provided in Table 2.2. It outlines the main benefits and challenges associated with textbook use. The challenges are most relevant to the discussion in the next section on how to move ahead with the textbook and improve on the challenges identified, most of which relate to the requirement for localisation.

Theme	Global Coursebook		Local Coursebook	
	Benefits	Challenges	Benefits	Challenges
▪ Basis in the local curriculum / ▪ Methodology ▪ Examination process	▪ Most often encourage a communicative language teaching approach ▪ Selection of activity types from which to pick and choose	▪ Not aligned to any one national/local curriculum ▪ Can be far-removed from local curriculum and examination system ▪ Can lack the full range of teaching methodologies espoused in the curriculum	▪ Curriculum-aligned ▪ Examination-oriented texts to help students achieve in national examinations	▪ Best language learning practices are sometimes not incorporated ▪ In some cases, over-emphasis on form-focused activities, comprehension and vocabulary learning ▪ Over-emphasis on examination-type activities ▪ Can lack the full range of teaching methodologies espoused in the curriculum
▪ Language level ▪ Language used ▪ Content	▪ Accompanying teacher manuals and resource banks can help teachers pedagogically ▪ Accompanying workbooks can help to focus follow-up tasks for reinforcement of language learned ▪ Accompanying materials can help English L2 teachers whose competency in English may be lower than desired ▪ Teachers can rely on the linguistic accuracy of the language presented ▪ Content is presented in a systematic way to allow for progression (e.g. recycling of language)	▪ Activity instructions can be too complex for teachers and/or students to comprehend ▪ The language used is sometimes difficult for learners to understand ▪ Communicative activities can be too difficult for teachers to support in class ▪ The length of texts and range of activities can put pressure on teachers in some countries, where they are expected to complete every page and activity of a coursebook	▪ The native language (L1) can be used for activity instructions ▪ Use of the L1 can support and scaffold learning where the level of English is challenging to learners ▪ Can support more grammar-oriented curriculums	▪ Sometimes the language presented can be ungrammatical or erroneous (basic errors in grammar and spelling) ▪ Can contain badly translated text from the native language into English ▪ Content can be presented in a disconnected manner with little recycling of language across chapters/topics/themes ▪ Can be lacking in content to progress all language skills equally ▪ Little uniformity in presentation of content ▪ They can provide inauthentic opportunities for learners to use English (e.g. two local students

Theme	Global Coursebook		Local Coursebook	
	Benefits	Challenges	Benefits	Challenges
	▪ Uniformity in how topics are presented and expanded upon ▪ Accompanying workbooks and answer keys can promote independent learning ▪ Texts are interesting with varying genres and a wide range of topics			asking for directions on-campus in English)
▪ Cultural acceptability ▪ Relevance to learner's life or experience		▪ English coursebooks are often specific to UK- and USA-based cultures and irrelevant to other cultures ▪ Can contain incorrect stereotypes (gender, country, language in use) ▪ Can be lacking in vocabulary which learners require to describe themselves or their community (e.g. rubber-tapper as a profession in Asia) ▪ Can contain activities which are at odds with local character (e.g. debate in polite Asian culture) ▪ Can be lacking in descriptions of local cultural activities or day to day life (e.g. night markets)	▪ Are specific to local culture / culturally relevant ▪ Can provide vocabulary to describe learners' local lives and experience ▪ Increased relevance for learners can result in better overall motivation and attainment	▪ Content can be too culturally-specific and become boring to the learner
▪ Production ▪ Quality	Visuals are attractive and motivating for students	Visuals can be distracting in some instances	Images can be more relevant to learners	Produced on tight budgets so visuals can be lower grade than global coursebooks

Theme	Global Coursebook		Local Coursebook	
	Benefits	Challenges	Benefits	Challenges
Cost		Can be prohibitive for parents in some countries	Are more affordable	

Table 2.2 Summary of benefits and challenges attributed to global and local coursebooks (compiled from sources in Tomlinson, 2008; OECD, 2014; INCA, 2014 and Ball and Feiman-Nemser, 1986)

It is evident that the push and pull surrounding textbook use in the classroom will most likely continue to be an ongoing debate. How is it possible to move ahead with the textbook?

2.2.4 Moving ahead with the textbook

Any negative impact on teaching and learning that is associated with the over-use of textbooks has a direct correlation to the quality of the textbook being used, how it was devised (e.g.Gottheim, 2010; Maley, 2011; Prowse, 2011), by whom it was written/developed (e.g. Bell and Gower, 2011; Harbison, 2008; Mares, 2003; Prowse, 2011, Tomlinson, 2008), its contents and how it was employed by teachers. Textbooks should be used in the same way as other classroom resources – they should be chosen for their relevance according to the teacher's course of planning on a daily/weekly/monthly basis, rather than dominating all planning, teaching and learning (e.g. DES 2005, 2005a, 2007; NCCA, 2008a; Bacha, Ghosn and McBeath, 2008).

The change which is emerging globally is a shift from global coursebooks to more localised ones (Arnold and Rixon, 2008; Prodromou and Mishan, 2008), which can account for local differences in culture (Lumala and Trabelski, 2008), language to be learned, curriculum and learners' familiar day to day lives (Dat, 2008).

In some jurisdictions (e.g. Greece), teacher competence to support learning using particular methodologies has created a need for more grammar-based textbooks, steering away from the dominant communicative-based ones (Prodromou and Mishan, 2008). These types of texts place teachers on a safer footing for predicting language which will be learned and encountered during lessons. In other countries (e.g. Japan), historical preferences for teaching methodologies persevere and teachers prefer shorter texts which are based on more behaviourist approaches to language teaching and learning, rather than communicative ones (Smiley and Masui, 2008). Other countries, where financial resources are an important consideration for textbook selection and use, rely on one core textbook which needs to be comprehensive enough to be the sole reference point for teaching and learning (Bolitho, 2008; INCA, 2013).

In moving ahead in Ireland, it is hoped that teachers will be able to divide out the combined tasks that textbooks have previously helped them to avoid (DES, 2005; 2005a, 2007; Gilleece et al, 2012; NCCA, 2008a), and carry them out more successfully (see Table 2.3 below). The two tasks – (1) planning teaching and learning and (2) sourcing relevant resources – can be and have been simplified. The NCCA have provided re-presented and

simplified overviews of the PSC, and an online planning tool to enable teachers to access the PSC more easily for planning (task 1). Similar assistance and guidelines for planning are provided by other organisations (for example, DES, NCCA, PDST and the National Induction Programme for Teachers (Primary), 2010/2011), among other educational bodies. However, these resources will not alleviate pulls on teacher time to any great degree or lack of teacher subject-knowledge, the latter of which could be alleviated through further teacher professional development. The requirement is for curriculum-based resources which are relevant to the local educational setting and which will provide them with more control to select and use materials which are relevant to the individual children in their classes.

Textbooks used to avoid		Re-addressed by
1	Curriculum-based planning – class and whole-school	▪ Re-presented curriculum ▪ Simplified curriculum overviews ▪ Online curriculum planning tool ▪ Planning guidelines from educational organisations
2	Sourcing relevant resources	▪ Curriculum-based source of content

Table 2.3 Summary of teacher requirements which have been addressed by textbooks

2.3 Consumable resources: digital/virtual resources

Having examined paper-based consumable resources, the discussion now turns to digital or virtual resources. The term *digital resources* encapsulates resources which are either online/web-based or offline. Examples of these types of resources include websites, PowerPoint presentations, electronic exercises and corpora. Digital resources covers a range of resource types, all with very different appearances, attributes and functions. It is more difficult to discuss digital resources in general terms than paper-based resources, as the digital entities offer so much variance between them. The following sections indicate general trends across the many digital resource types.

2.3.1 Integration of digital resources

Paper-based resources require literacy for their content to be used and accessed. Digital resources, on the other hand, require a certain level of ICT infrastructure (see Section 2.4 below) and ICT expertise of the teacher/practitioner and/or learner. These factors have slowed down the uptake of digital resource use and integration into day to day teaching and learning (DES, 2008; European Commission, 2013; NCCA, 2007; Shiel and O'Flaherty, 2006). Further detail on these reported inhibitors to ICT use and integration include a lack

of reliable equipment, teacher confidence, teacher competence, teacher resistance to change, national policy and school policy (European Commission, 2013; NCCA, 2005a, IPPN, 2008). The European Commission (2013) reported that it was pedagogical knowledge and subject knowledge of ICT integration (the availability of subject-relevant tools and how to integrate them into teaching and learning) which were the greatest contributors to higher ICT integration in classrooms, rather than teachers' personal ICT ability. The specific reasons cited by teachers in Ireland for their low uptake of ICT are outlined in Table 2.4 below.

Reason cited	**Expansion of reason**
Time constraints	▪ teachers report to already be stretched for time ▪ integrating ICT takes a lot of time ▪ lack of time for planning and finding resources ▪ level of support required per child ▪ use of ICT at a cost to other subjects
Unsuitable/lack of software and resources	▪ unfamiliarity with suitable resources ▪ lack of computers and/or internet access ▪ lack of subject-specific and relevant resources
Lack of classroom management strategies for ICT use	▪ class size (child to computer ratio of 30:1) ▪ multi-grade setting (one computer available for all children) ▪ discipline (difficult to retain control)
Insufficient numbers of computers available/accessible	▪ one computer in the classroom ▪ lack of space for the entire class in the ICT/computer room, sometimes only 15 computers for the entire class
Lack of teacher confidence or skill in ICT	

Table 2.4 Teachers' reported reasons for low levels of ICT integration (compiled from DES, 2009; NCCA, 2005a, 2008a).

In Ireland, teachers showed a preference for CD-ROM based software over online subscriptions (Shiel and O'Flaherty, 2006; NCCA, 2008a), potentially due to their reliability and increased teacher control over their subscription and/or ownership/access. However, teachers also reported a lack of relevant subject-specific ICT resources (DES, 2005, 2008). Children's use of ICT was often reported to be restricted to the use of word processing software (NCCA, 2005a, 2008a; Butler, Leahy, Shiel and Cosgrove, 2013). Similarly, across Europe it was noted that "students' ICT use during lessons still lags far behind their use of ICT out of school, affecting their confidence in their digital competences" (European Commission, 2013, p. 16). These findings contradict the many recent Irish policy documents relating to increasing and embedding student use of ICT in educational reform (e.g. Project Maths, Junior Cycle Reform, Key Skills Framework) (Butler, et al, 2013).

The majority of teachers have been reported to use ICT to prepare their own lessons, with an average of 29% reporting to use ICT during the course of their lesson (European Commission, 2013). In Ireland, this figure was at 75% (European Commission, 2013).
Two of the more recent Irish strategy documents relating to ICT in education[2] (DES, 2000 and DES and ICT Ireland, 2009) have both set out the availability of curriculum-based digital teaching and learning resources as one measure required for the successful integration of ICT in education. Coupled with the ongoing drives for more ICT facilities in schools and greater ICT integration in teaching and learning, digital/electronic resources are certainly required and warranted. Indeed, some publishing companies have recognised the need to fill the paper-based and virtual/electronic requirement of teaching and learning through their investment in *learning packets*, that is, paper-based resources which have complementary online/virtual/mobile learning elements. It is certainly necessary to ensure teachers are equipped for all eventualities (ICT availability or lack thereof) and their levels of ICT ability; both paper-based and virtual/digital resources are required.

2.3.2 Samples of digital resources

Adding digital capabilities to any educational setting can "allow information and resources to be created, accessed and shared on virtually every subject matter imaginable" (DCENR, 2013, p. 22). Taking the first of the three functions of creating, accessing and sharing, two examples of the ***creation*** of digital resources are teacher's own MS PowerPoint slides and the creation of Hot Potatoes exercises. PowerPoint allows teachers to develop a set of slides to present information to their learners. These slides can be enhanced with animation (moving elements of the slide content), images, web links, embedded or linked videos and audio. PowerPoint slides are reusable, can be shared among other teachers and are easily edited to allow them to be localised or further customised to their target learners. Similarly, learners can use PowerPoint to prepare assignments or schoolwork, like presentations, project information or visual displays.

Hot Potatoes[3] is a free downloadable authoring tool which allows six different types of exercise (e.g. Cloze tests, multiple-choice exercises, click and drag exercises) to be developed for use online or offline. It allows the teacher/practitioner to customise these exercise-types with their own content. Teachers can also predict typical errors their learners may make and supply feedback on each one. Once created, the finished exercise can be accessed by learners online, or offline. The exercise creation file can also be edited later,

[2] At the time of writing in 2015, the report on the ICT Census 2013 (DES, in process) remained unavailable.
[3] https://hotpot.uvic.ca/

to amend a previously created exercise, hence allowing further localisation at a later date. There is not any limit on the number of times an exercise can be completed. Hot Potatoes is still very much in use among practitioners, despite its developers no longer developing or supporting it.

The second function, ***accessing***, is exampled by the website IXL[4] and the CD ROM series of EuroTalk Interactive language learning software. IXL is a global platform which has been localised to many national curricula. It provides curriculum-based Mathematical practice for students. The website description explains how the website can be used to help students to practise curriculum-based Mathematical areas (e.g. Geometry, Statistics) and prepare for standardised testing. The platform monitors students for teachers' later access. The IXL platform is pre-formed, in that it already contains a bank of exercises for learners to complete, which cannot be edited or further localised by teachers. These exercises are computer-generated which results in there being a limitless supply of exercises for practising any one Mathematical area. Where content is pre-formed, it does not allow for editing, further localisation or reuse outside of the platform. However, some amount of reuse may already be evident within the platform where the listing of Mathematical areas is finite among all national curriculums covered. This would mean that exercises can be recycled across the different localisations available.

EuroTalk Interactive is a series of language learning CD ROMs which cater to learners with varying levels of language competence. The beginner's offering (Vocab Builder) is a topic-based picture dictionary. It presents text, image and audio pronunciation through presentation of the vocabulary. Games are used for practising learned vocabulary. The learner can record themselves for oral practice and compare their own voice pronouncing phrases or sentences to the same pre-recorded utterances. As pre-formed content on a CD ROM, no editing or localisation can be carried out on the content by a teacher/user. However, the content was reusable by the CD ROM creators, as they offer the exact same content and layout across many languages being learned. A translation of their language file, along with recreating the associated audio files would result in a new language being developed.

The second and third of the three functions, ***accessing and sharing***, could be exampled by a myriad of online websites and resources. Two resource-focused websites which allow access to and sharing of resources are Scoilnet[5] and TES Connect[6]. Scoilnet is an Irish

[4] www.ie.ixl.com
[5] www.scoilnet.ie
[6] www.tes.co.uk

website which acts as a portal for subject-specific websites (e.g. www.german.ie), nationally-available educational resources (e.g. Britannica School) and the access and sharing of teachers' resources. The website hosts over 13,000 resources (DCENR, 2013) which are freely available to download.

TES Connect is similar to Scoilnet except that it is a UK website which is localised to the curriculum in the UK. It is a hub for sharing and downloading teaching resources. The resource area is a little more specific than on Scoilnet in that varying Creative Commons Licences are applied to the resources there. These help to protect the resources to varying degrees, such as the original author being cited when used and permissions associated with editing and re-sharing the resource. Teachers can also charge for use of any of their resources, or pay to use others' resources. In 2012, over 1.18 million resources were downloaded from TES Connect by Irish practitioners (DCENR, 2013, p. 28). Some of the content on both Scoilnet and TES Connect can be edited and further localised (depending on the resource type). Reuse of content depends on its creative commons licence (as per TES Connect) or the format of the resource chosen (e.g. the content from a PowerPoint file could be easily accessed and reused whereas other file-types may not facilitate this).

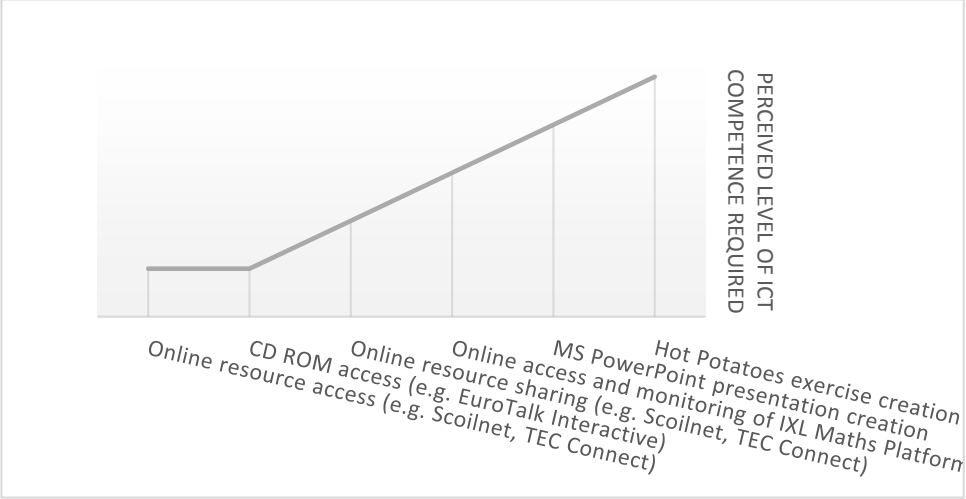


Figure 2.2 Perceived level of ICT competence required for each of the digital resources discussed

As seen above, the level of teacher competence in ICT can be an inhibiting factor in the use and integration of digital resources. Figure 2.2 illustrates the perceived increasing level of ICT competence required by teachers to access and use the sample resources outlined above. Accessing resources (downloading or viewing) can be perceived as being the activity requiring the least amount of ICT competence. More competence could be needed for sharing resources online (uploading). Accessing and monitoring require an even greater level of ICT competence, with the creation of resources being perceived as requiring the

greatest level of ICT competence. The level of interaction the teacher has with the digital resource (to use, integrate, edit or localise the resource) is also a determining factor in the proportion of ICT competence required. An increased level of interaction will require an increased level of ICT competence and vice versa.

2.3.3 Control, supply and selection of digital resources

Paper-based resources need to be selected (potentially from prescribed lists), examined and purchased for use in the classroom. The selection and examination of digital resources follows suit. There are many digital resource guidelines which offer teachers and practitioners support in the selection of suitable digital resources (e.g. Education Services Australia, 2012 for evaluating digital resources for education; Hubbard, 2006 for CALL evaluation; Leacock and Nesbit, 2007 for evaluating multimedia learning objects; Davies, 2011 for the evaluation of CALL software and websites). Offline digital resources also need to be purchased, along with some online resources requiring a subscription fee. However, the majority of online digital resources are free to access and use and are generally easier to obtain and access than paper-based ones (where required ICT peripherals are in place). The world wide web presents innumerable opportunities to access authentic language, ready-made resources and teaching inspiration for teaching and learning languages. Similarly, it provides opportunities to communicate with other learners and for all curriculum-subject, there are many opportunities to access ready-made resources (as with the Scoilnet and TES Connect examples above) or gather together a selection of content to create one's own resource. The prevalence of Open Educational Resources (OERs), which are resources which are freely available for educational and research purposes (such as those found at Scoilnet and some at TEC Connect as described in Section 2.3.2), also add to the myriad of freely available resources.

Using ready-made digital resources, particularly online resources, requires a high level of critical thinking and analysis to ascertain whether they are of suitable quality, relevance to the learning situation or suitable for the learners. Resources can be shared by anyone online, and the author background or expertise can be unknown. Searching the web for relevant resources using a search engine can potentially start a long 'search and explore' process for teachers and practitioners. Websites containing educational resources (e.g. Scoilnet) reduce this search time to a degree by allowing educational resources to be accessed by relevant criteria (such as educational level, subject and/or curriculum area) but it too may also require much time (e.g. Greene, 2013 outlined how a lack of tagging on online resources inhibited their easy access).

2.3.4 Moving ahead with digital resources

The level and amount of digital resources available is increasing daily. Digital resources offer an enormous range of possibilities for teaching and learning through their creation, access, sharing and through their use. They create opportunities for more interactive learning, where learners can also communicate with one another.

The challenges associated with digital resource use are the requirements for ICT peripherals (structural and physical resources) as well as suitable ICT competence. The advantages include the positive learning opportunities they offer, the ability to more readily edit, localise and reuse certain types of digital resources.

As mentioned in the introduction to this section, there is a lot of variation in digital file types which impacts the ability to localise (or level of localisation) and reuse them. Section 2.5.2 below examines the localisation of digital resources.

Moving ahead with digital resources is somewhat dependent on the challenges listed above. The next section examines the status of the structural and physical resources required to support them.

2.4 Structural and physical resources: ICT

There is a requirement for certain structural and physical ICT resources to support the use and integration of consumable ones. The examples provided previously were the need for computers to mediate electronic exercises and to allow software CD-ROM or DVDs to be used or watched. This section briefly examines the levels of ICT penetration in educational settings globally and locally in Ireland. While consumable paper-based resources prevail and dominate in educational settings, it is also important to future-proof any resource development to ensure that it can be easily embedded in more ICT-enabled schools, if or when they come on stream to the required level.

Globally, student to computer levels were reported to be at reasonable levels in most European countries (between three and seven children per computer), with more computers available at secondary level than primary (European Commission, 2013). Huge variation existed across the countries examined, with internet connectivity rates varying from 2mbps to 100mbps and the ratio of students to internet-connected laptop varying from 10 to 250 (European Commission, 2013). The majority of computers were available in computer rooms, with this trend increasing in upper education levels (European Commission, 2013).

In Ireland, levels of ICT resources have improved over the years, with current levels at 6:1 and aligning closely to the EU average (European Commission, 2013). Most primary schools were reported to have an average of one computer in each classroom with broadband access (Gilleece, Shiel, Clerkin and Millar, 2012; DES, 2008; NCCA, 2007; Shiel and O'Flaherty, 2006). However, Ireland still lags behind on internet and school network connectivity (Shiel and O'Flaherty, 2006). The student to computer ratio masks some issues surrounding the age and reliability of computers in schools (DES, 2008; NCCA, 2007; Shiel and O'Flaherty, 2006). Similarly, the financial burden of maintaining existing equipment delays further progress with addition to existing equipment (DES, 2008; Shiel and O'Flaherty, 2006). The European Commission (2013) reported that it is a mixture of school-staff, external companies and educational authorities who are tasked with maintaining ICT equipment in schools across Europe. The trend is for in-school staff to take charge of this work (European Commission, 2013) with support from external companies. Ireland had one of the highest figures for outsourcing this work (European Commission, 2013).

One of a series of recommendations made to progress ICT provisions in schools noted that "Innovative, high-quality and Irish curriculum-related digital teaching and learning resources must be made available to teachers and students at primary and post-primary levels. This is a core principle in the successful integration of ICT in schools" (DES and ICT Ireland, 2009, p. 19).

2.5 Localisation of consumable resources

When referring to paper-based resources, localisation is the process of customising the resource (e.g. textbook) to its target users. This could involve cultural adaptations, linguistic ones and/or changes to the pedagogical focus (e.g. a particular curriculum). The discussion surrounding the localisation of paper-based resources stems largely from the textbook industry where global coursebooks dominate in minority settings.

When examining digital resources, the same issues emerge for making content more relevant to learners. Adapting digital content brings with it an additional level of complexity in some cases, where content can be embedded in a software application or website. This moves the digital localisation discussion closely in line with the definition from the software localisation industry, with one added caveat, that the pedagogical and curricular context also need to be considered. "Localisation is the process of adapting digital content to culture, locale and linguistic environment" (van Genabith, 2009, p. 4).

Van Genabith (2009) outlined that personalisation is one of the challenges in the localisation industry, where "information" is the entity to be localised (p. 5). Van Genabith posited that "information is most useful if adapted to personal requirements, backgrounds and goals of a user" (p. 7) with the person being "the ultimate locale" (p. 5).

2.5.1 Paper-based resources

As seen in Section 2.2.1, some textbooks are produced at a national level which ensures that they are aligned to the national curriculum. This means that the nationally produced textbooks are a basic resource in all schools in those countries. In other jurisdictions, textbooks are produced commercially by publishing houses (global or local), and may or may not adhere to a curriculum or curriculum-led guidelines. Global textbooks are sometimes referred to as "imported coursebooks or foreign coursebooks" (Dat, 2008, p. 265), which are geared towards a large audience and are often culturally irrelevant to the majority of its audience. Local coursebooks can be referred to as "in-country or domestic coursebooks" (Dat, 2008, p. 266) and are produced with more relevant local cultural insight.

A further form of the coursebook, the *localised coursebook*, is present in some systems and is a half-way point between global and local coursebooks (Figure 2.3). This is a global coursebook which has been localised to the target environment (with or without copyright issues taken into account) (Barrios, de Debat and Tavella, 2008; Dat, 2008). Arnold and Rixon (2008) referred to these as *glocal* materials (Arnold and Rixon, 2008, p. 40), that is, global materials which are adapted by local writers in response. They are usually developed in response to a new initiative for teaching language in a country that leads to a short-term gap in the market for locally-produced materials. Glocal materials then give way to locally produced materials which overtake the need for imported books and their adaptation.

Global textbooks are produced by large publishing houses and have a far reach in their audience. In particular, EFL or ESOL classrooms around the world are dominated by a small set of well-known resources. Global textbooks have been cited to lack cultural relevance, often being far-removed from the day to day lives of the teachers and students who use them (Dat, 2008; Prodromou and Mishan, 2008). They can also present a picture of everyday life, which is at odds with the day to day existence of students (Dat, 2008). In this way, they can be lacking in relevant vocabulary to facilitate students in talking about themselves, their day to day activities and experiences. Prodromou and Mishan (2008) similarly noted a need for more localised materials which meet the local need match the cultural backgrounds of the learners, the language background of learners and their interests

(p. 208). Dat (2008) provided many specific examples of cultural misalignment of global textbooks (Table 2.2), one in particular highlighted how Thai students were reported to dislike debate as a form of communicative activity, as Thai culture encourages "agreeable, harmonious exchange of ideas" (p. 271).

In some cases, practitioners are working in a curriculum void (Lumala and Trabelski, 2008). As a result, they are required to patch together and localise materials from a selection of those which are available to them. Lumala and Trabelski (2008) reported on their process of providing worksheets around any one topic which were sourced from in-house resources such as the university library, the internet, newspapers and magazines. This method is a little ad-hoc offering little continuity or congruence across the materials presented, but does add variety for students.

In other scenarios, the global textbook industry can start to miss out on local market potential due to their mainstream-focused texts. As a result, the publishing houses have been proactive in producing locally relevant texts to accompany their textbooks. Greece provides an example of this, where local grammar-based and vocabulary texts were produced to accompany the global textbooks. These sources were required by local English L2 teachers as the use of global textbooks with a communicative ethos was reported to often place a strain on Greek teachers' linguistic abilities (Prodromou and Mishan, 2008). Teachers would be required to allow impromptu conversation which they may or may not be able to support for error correction. Teachers were more readily able to support a grammar-based programme, as it can be "much easier for a teacher who has learned English from books and tapes to teach grammar than to teach the pragmatics of English" (Prodromou and Mishan, 2008, p. 203). Global publishers soon followed local publishers in publishing supplementary grammar- and vocabulary-based text in order to adapt and localise to the market demand. Later publications were reported to be "a hybrid constructed from the UK tradition of textbook design and the local penchant for lots of controlled practice" (Prodromou and Mishan, 2008, p. 203).

Sometimes, a new educational process, such as standardised testing, can call for textbook content to be *localised to the test* to maximise student achievement. As standardised testing has become the norm across the USA, most publications are meeting teacher demand to produce more grammar-oriented texts (Frazier and Juza, 2008), which stray from prevalent communicative-based curriculums in place. Standardised testing is also becoming more prevalent across Europe (OECD, 2014), with similar effect on textbook materials. Prodromou and Mishan (2008) referred to this process as "washback", defining it as "the effect, direct or indirect, of examinations on teaching methods and materials. Washback

may be positive or negative and its influence is greater the more important the examination is" (Prodromou and Mishan, 2008, p. 202).

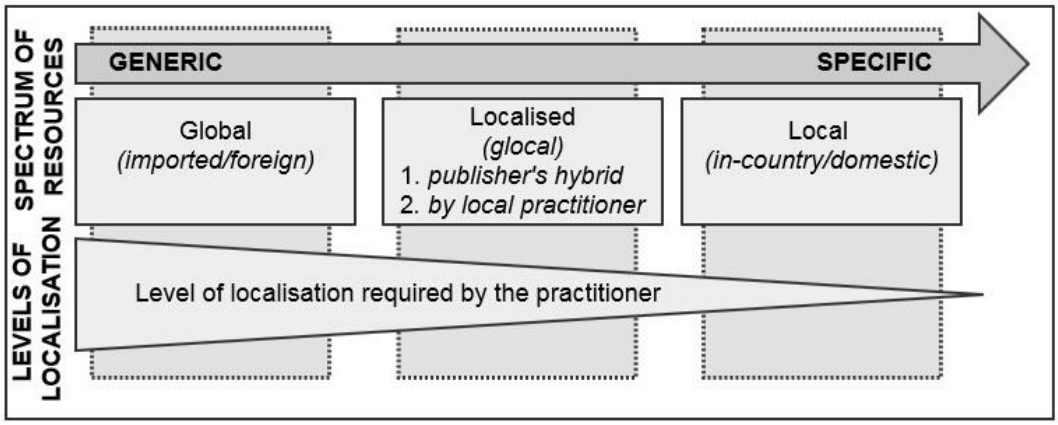


Figure 2.3 Varying levels of localisation required, by resource type

One extreme in the localising process, is where the education system localises itself against the benchmark of widely available textbooks. This is done to ensure the availability of quality resources. The USA provides an example of this where many publishing houses align their texts to the "big three … adoption states of Florida, California and Texas" (Frazier and Juza, 2008, p. 179) as they form a large proportion of market sales. As a result, smaller states tend to match their standards to one of these states to ensure they can readily access textbook materials (Frazier and Juza, 2008). They localised their education system, rather than the textbook.

The requirement to localise global textbooks is a lesser one in primary and secondary school settings, where subject-specific textbooks (local textbooks) are required to cover the contents of an established curriculum and are considered to be local textbooks (as illustrated in Figure 2.3).[7] A further level of localisation is also required and apparent in these educational settings where the practitioner/teacher localises (or customises) these learning resources for their particular learning group. Maley (2011) described this requirement as a result of there "never be[ing] a perfect fit" (p. 379) between the materials, teacher and learner despite instances where materials are even designed for a very specific target group. He describes a series of "coping strategies" for teachers to use to adapt materials for their learners such as "omission … addition … replacement … reordering …" (p. 381-2). A second approach Maley (2011) recommends to try to overcome the need to "bridge the gap" between materials and learners is to change how materials are developed, to allow for "greater flexibility in decisions about content, order, pace and procedures" (p.

[7] The discussion in Chapter 3 (Section 3.2.1) will address how the interpretations of the curriculum can become skewed by the many agents and processes (e.g. teacher, publisher) which interpret the curriculum before it reaches the student in the classroom via a textbook.

380). If working with younger learners in a primary school setting, there would be a need for materials at "appropriate cognitive and maturity level for the age of the children" (Arnold and Rixon, 2008, p. 46). These localisation processes are generally carried out by the teacher through sourcing, editing and adapting a variety of resources and supplementary materials, or indeed adapting the contents of the local textbook.

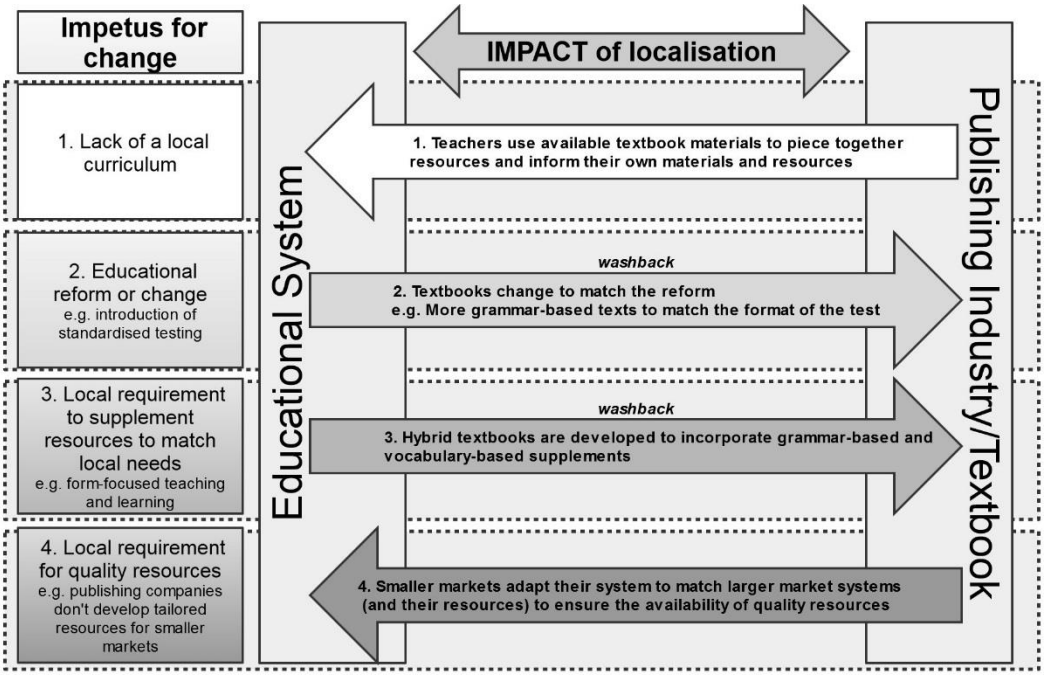


Figure 2.4 Direction of system impact of localisation with varying impetuses for change (compiled from references in this chapter)

It is apparent, that no published text will ever completely meet the needs of its target users. Varying levels of localisation are required in all circumstances. The impact of this localisation usually transfers from the specific educational system to the textbook (washback) with the curriculum or an educational initiative initiating the change. In some circumstances, the impact can also transfer in the reverse order with the textbook influencing the educational system, as seen in the smaller states of the USA. Figure 2.4 above illustrates this phenomenon.

2.5.2 Localising digital resources

The global ethos of the world wide web has created a greater tolerance for multicultural contexts in online resources. However, learner preference can still remain for more culturally specific digital resources (e.g. Shen, Yuan and Ewing, 2015). As mentioned above, digital resources encompass a wide range of file types, platforms and resource types, so the need to localise as well as the localisation processes for each can vary greatly. Ryan, Anastasiou and Cleary (2009) provided a definition of digital content which included any

content "published as a computer file" (p. 12) and which is stored on an electronic device. Their taxonomy included digital content ranging from personal emails, to online help and YouTube videos to SMS messages.

With the localisation of digital resources falling more in line with the processes employed to localise software, certain applications have been developed for these purposes. Translation memory software (computer-aided translation), software localisation suites and machine translation (Ryan, et al, 2009) have all supported the assisted translation and automatic translation of content or information. Ryan, et al (2009) outlined how companies have tried to reduce their localisation costs by establishing "content development guidelines, best practices and standards" as well as the "internalisation of content" (p. 18). Internalisation is the "process of generalising a product or document so that it can handle multiple languages and cultural conventions without the need for redesign" (Ryan, et al, 2009, p. 19). Of note, is that the processes involved in internalisation within the localisation industry are concerned with abstracting the content or information away from its surrounding product or system, to allow it to be translated into other languages. An attempt is similarly made to confine culturally specific elements to accessible areas which can be easily altered.

Ryan, et al (2009) outlined the varying processes within the internalisation process. Figure 2.5 illustrates these processes. The initial *authoring* of content can be confined by controlled language rules and lexica. Controlled languages result in a simpler language output, thus aiding readability and easier translation. Enabling in software development means that the software is developed with the ability to be translated into multiple languages. Content storage and access are considered and content is kept separate from the software processes and functions of the system under development. Testing ensures that the authoring and enabling processes were conducted effectively and that the overall system is tested for functionality.

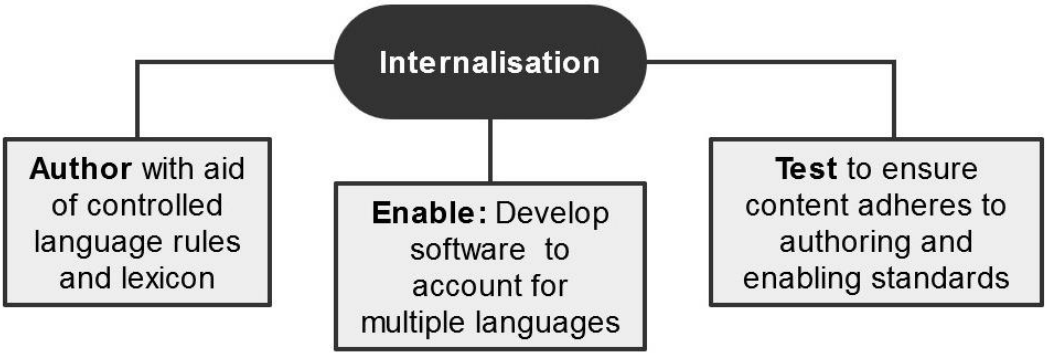


Figure 2.5 The internalisation process (compiled from Ryan, et al, 2009)

A comparison can be made between the internalisation process and the development of educational resources. Both processes are involved with developing digital content from scratch. Internalisation accounts for future localisation needs by making sure that the product under development is 'suitable' for localisation. The development of educational resources could be enhanced by carrying out a similar process during the development stage, rather than being applied retrospectively. Figure 2.6 outlines how these two processes map to one another.

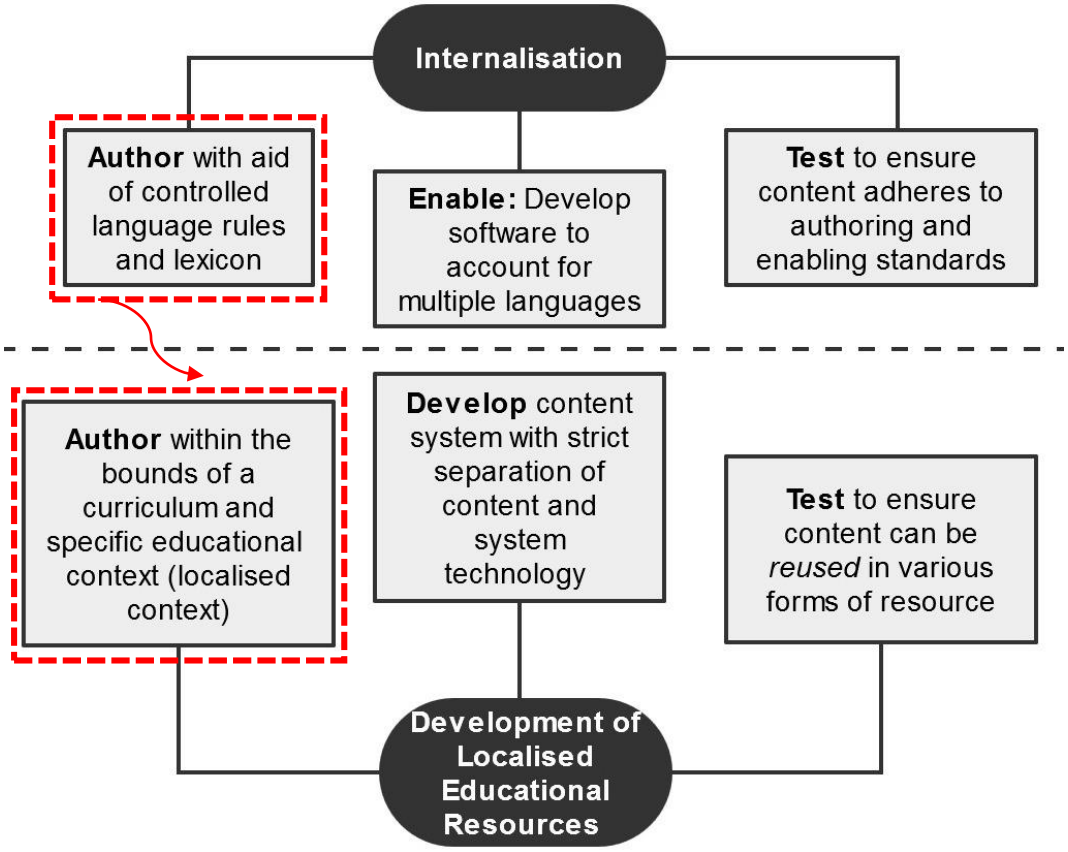


Figure 2.6 Mapping the development of localised educational resources to the internalisation process

A notable link can be made between the processes of internalisation authoring and the development of localised educational resources authoring (Figure 2.6 connections drawn in red). While the former focuses on limiting the content through the use of a controlled language, the latter ring-fencing of content is made by localised contextual settings like the curriculum. The discussion will return to examining the parallels between the two types of content (internalised and educational) to examine whether the localised educational language can also be deemed a controlled or ring-fenced language. The content analysis processes of Chapter 6 explore this parallel further in order to make conclusions about the characteristics of the educational content.

A final comparison can be made between the **retrospective processes** involved in the localisation of educational resources and localisation within the translation industry (rather than internalised localisation); this time it is a dissimilarity. The two processes have much in common, but the localisation of educational resources must also account for the local curriculum, established pedagogy, learner requirements and contextual variables. Rather than a translation of content, a change and/or addition to content may also be necessary to account for these factors. However, localisation within the two domains is largely concerned with isolating the content which needs to be altered (localised) from its surrounding function or digital product. The discussion will return to the abstraction of content in Section 3.1 (p. 48 Defining raw linguistic content).

The need to localise consumable resources for their required setting is a very relevant and important endeavour for teachers and practitioners, as well as being an important consideration for resource creators. The next chapter examines how consumable resources have been and can be created.

2.6 Summary and conclusion

This chapter provided evidence of the use of and, benefits and challenges surrounding educational resources. Regardless of which countries are investigated, resources in one format or other have been shown to form the basis of all teaching and learning. A categorisation of educational resources was provided which grouped them according to their function and type. One type, consumable paper-based resources (e.g. textbooks and coursebooks), are amongst the most prevalent used in primary and secondary schools as well as other educational settings. Consumable digital resources are increasingly being used, but their integration into educational settings is dependent on teacher competence in ICT and access to the structural and physical ICT resources needed to support them. Despite some settings attempting to integrate consumable virtual/electronic materials, levels of ICT infrastructure to support these electronic/virtual formats of resources are not always at the required level.

A recurring set of themes emerged which highlight the main issues discussed in respect of textbooks. These themes reflect some of the section headings within this chapter:

- Textbooks are a core and staple resource in educational settings. Digital resources are widely accessed but mostly used by the teacher to prepare and present lessons, and only occasionally by their students/learners.
- Teachers have been shown to have an overreliance on textbooks for their planning, teaching and learning, due to a lack of subject knowledge, a lack of time and to replace

the need for planning from curriculum documents. Digital resources are used to a much lesser degree, but present a myriad of options for interaction, communication, and creating, sharing and accessing resources.

Section 2.5 outlined how all types of textbooks and digital resources (global, localised or local) require varying levels of localisation to meet the needs of their local consumers. The more generic the resource, the greater the level of localisation required, and the more specific, the lesser the level of localisation required. The impetuses for localisation on a resource can stem from a lack of local curriculum or educational reform for example, but the localisation process has the potential to impact the educational system or the publishing industry, depending on where the impetus began. A new summary of localisation impetus and impact was proposed in this regard (Figure 2.4). The localisation industry employs an internalisation process whereby new digital resources under development are made 'localisation-ready' from their build point. The same principle could be applied to the development of educational resources. This same section on localising digital resources also presented two key themes to the research – that of *abstracting content* from its product and the parallels that exist between internalised authored content developed within the confines of a controlled language and the *controlling or ring-fencing of content in educational resource generation* to features of its local context (e.g. the curriculum).

This chapter has mapped out the requirements for educational resources –

- the separation of content from its product/technology
- reusability and transferability of content to maximise the time/cost/expertise invested in their original production,
- localised and curriculum-based educational resources.

The next chapter examines research which has separated educational content (in the form of databases) from its product. It also explores how consumable educational resources are created, developed and analysed and formalises the research space for resource creation.

Chapter 3 The development and analysis of educational content and consumable resources

This chapter begins to pull all of the disparate elements of this research together. Following the Chapter 2 discussion on resources and the challenges associated with consumable resources in educational settings, the discussion now focuses on how educational resources are developed. It examines whether any existing development processes or content systems address the three challenges emerging for consumable resources: (1) separation of content from its product, (2) reusability and transferability of content and (3) localised and context-specific resources. It also specifically examines the place of 'content' in these development processes, that is, the teaching and learning content, as separated out or abstracted from its product (be it a textbook, CALL courseware or corpus-based teaching resource). The purpose of examining these principles and methodologies for relevance to the three challenges and content creation is to begin the process of piecing together a solution to the challenges outlined.

A parallel thread weaves into the discussion in this chapter - that of examining (analysing) the make-up and characteristics of educational content. The importance of this process lies in exploring how educational content which has been abstracted from its product/resource-form compares with more traditional types of educational resources (e.g. textbook or authentic texts from newspapers) and whether its form and characteristics have further reaching implications for its use and manipulation.

Section 3.1 initially defines *raw linguistic content* as the research term for abstracted content. Section 3.2 examines how consumable paper-based resources are created and what principles and methodologies are at play in their construction and development. As outlined above, the focus of the discussion is on whether resources created using these processes would address the three challenges identified for resources, and how important the focus on content generation is. Section 3.2.5 examines content systems and databases which have made progress with one of the resource challenges - the separation of content from its surrounding product - with the aim of creating educational resources from that content. With much effort to be expended in the development of localised raw linguistic content, an examination of this content becomes important in ascertaining how it differs to other types of educational resources. Section 3.3 examines how resource content can be analysed and characterised. This chapter concludes by drawing all of the emerging issues and findings from the discussion in Chapters 2 and 3 together, to form the global problem space for the research outlined in this thesis.

3.1 Defining raw linguistic content: abstracted content

The term *raw linguistic content* was first introduced within Section 1.4 in Chapter 1. It was introduced as a concept that may not sit easily with language teaching practitioners, SLA theorists or CALL researchers as it temporarily disassociates language learning content from its associated didactics. The teacher or practitioner then selects relevant teaching content and reintroduces the didactics as the teaching professional. The term was also referred to in Section 2.5.2 in Chapter 2 where the discussion surrounded the localisation of digital resources. A similar process of content or language abstraction is carried out in the localisation industry to enable digital resources or software to be localised for new contexts. Once this content has been translated and cultural adaptations have been made, the content is re-associated with its product to form a localised whole once again. The discussion below expands on these introductions and encounters, to place raw linguistic content in its theoretical background.

Maley (2011) proposed the notion of 'flexi-materials.' These materials are based on Prabhu's (1988) idea of semi- and meta-materials. These latter notions provide choice for the practitioner and learner, in providing a range of relevant teaching content and activities, from which they can choose those which are most relevant. In doing so, they are placing the content choice, order of presentation, pace of presentation and procedure in the practitioner's or learner's control (Maley, 2011).

These notions of choice in content and materials can be taken one step further. In focusing in on the **written** *content* of consumable resources or materials, the aim is to look beyond the presentation of materials in texts and prepared materials and focus in on the *words, phrases and sentences* embedded there. When materials are being used directly in teaching and learning the content is generally embedded in the end product, be it CALL courseware, a textbook or a vocabulary SMS task (such as a pairing or odd one out activity as reported in Levy and Kennedy, 1999; or a multiple-choice item as reported in Keogh, 2011). In short, whatever format content takes in language learning materials, the *raw linguistic content* always comes back to the building blocks of language – words, phrases and sentences – which are embedded in the end product.

The use of any resource in an educational setting operates in a similar way to raw linguistic resources, by presenting content to be learned and practised.

The importance of content in its 'raw' or bare format is that it forms the basis for all that is learned and taught through resources. Decoo (2011) describes the importance of 'words' as "the tangible and quantifiable units that make up language. Anyone who starts to learn a language will do so by hearing or reading words…" (p. 102). In focusing on content, all other language learning factors (e.g. student motivation, student learning style, technology to be used, etc.) are stripped away to be left in the domain of the teaching practitioner. Bare content is exclusive of all of these other extraneous elements. Student- and context-specific details such as typical student errors, known languages and national curriculum (named localised settings in this research) can be added to this raw language, to localise it further for its particular language learning environment. The other factors, such as student learning style, motivation and output type to be used (e.g. paper-based, electronic exercise or mobile phone SMS) can be decided later and incorporated by the teacher, who is best placed to match the content to a suitable activity for his/her students. Raw linguistic content is reusable, in any setting which is teaching the same language to the same level. An older learner may require a higher register than a younger learner, but the basic content for both can be used in both instances.

Through an SLA lens, this approach may seem more structuralist and cognitivist in its stance, where 'content' to be learned and mastered is delimited and centred around language as a 'code' to be learned. Structural or linguistic approaches (e.g. Bongers, 1947; Palmer, 1917) examine the underlying processes of language learning. These approaches account for L2 learning through examining the rules which govern the structure of a learner's L2 knowledge through processes such as learner error analysis (see Section 4.1.4) and contrastive analysis. Some cognitivist SLA theories describe language learning as "the acquisition of constructions that map linguistic form and function" (Ellis, 2006, p. 100) and embrace that the language learning process is aligned with other types of human learning. Constructs are used as the learning unit and can comprise words, phrases and/or sentences which are learned through "using language and engaging in communication" (Ellis, 2006, p. 101). Cognitivist approaches are based on examining and explaining how the language input and order of learning processes impact L2 learning. Pienemann (2008) proposed one such theory, called processability theory, which outlined the processes and stages through which an L2 learner progresses as they acquire a language. Pienemann (2008) posited that a learner could only progress through from the earlier stages of memorisation of single words to more complex embedded subordinate clauses for example, as his/her brain adapted and learned from the earlier and simpler processes.

Given all of the main approaches which can form the basis of work in SLA and the teaching methods which can stem from these approaches, it will always ultimately fall to the teacher to formulate their own viewpoint on how the relevant approaches and methods outlined in the curriculum will be interpreted through their pedagogy in the classroom. Raw linguistic content

interacts with the scenario of teacher control and choice by forming a basis in which they can ground their teaching and learning resources.

Raw linguistic content that is based on a curriculum

The raw linguistic content relevant to a curriculum can be generated from the curriculum details. For example, a language curriculum might contain themes and topics around which language should be learned. One example (illustrated in Figure 3.1 below) using the Primary School Curriculum in Ireland shows how a curriculum subject comprises many items, one of which is the core content to be taught and learned. The dotted line connecting school/classroom planning to the curriculum shows how school and classroom planning can define a curriculum in a school and classroom, but they are not a component part of a curriculum statement.

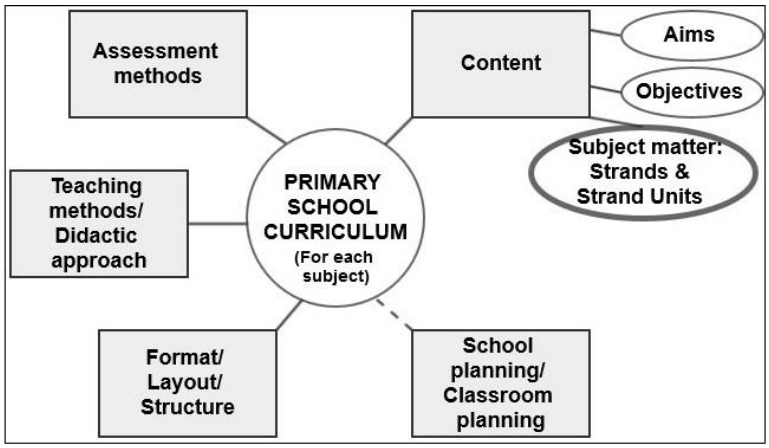


Figure 3.1 Narrowing in on subject matter for the production of curriculum-based raw linguistic content

Raw linguistic content from existing published resources

Raw linguistic content can be similarly extrapolated from its home or position in published resources (e.g. textbook or online exercise). However, this extraction of raw linguistic content is a time consuming endeavour, as the raw linguistic content is generally mixed within details of the text layout, image positioning or programming language, mark-up language or script used (Figure 3.2). Through extracting raw linguistic content, content is disassociated from its didactic form in teaching and learning. The raw linguistic content can be returned to its didactic form by the teacher or practitioner, when selected and inserted into a usable template (to produce a paper-based resource, or a virtual resource, for example) which converts and/or returns the material to a usable teaching and learning resource.

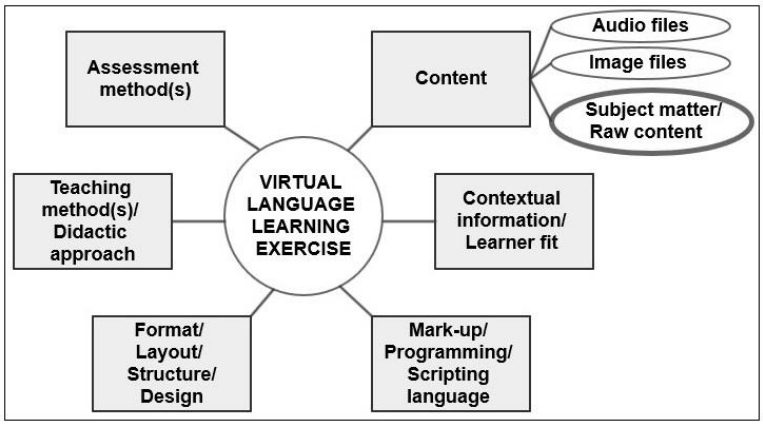


Figure 3.2 Isolating subject matter or raw linguistic content from virtual language learning exercise

3.2 The development of teaching and learning resources: established principles and methodologies

The first part of this section examines general textbook writing and publishing norms, with particular emphasis on how these processes interact with the messages inherent in a curriculum. Similar emphasis is also placed on how each of these processes considers content in their development. Sections 3.2.2 and 3.2.3 describe the existing and prevailing development principles, methodologies and frameworks from within (1) the Language Materials Development discipline and (2) the Computer-Assisted Language Learning (CALL) domain. Section 3.2.4 discusses the use of corpora, or electronic collections of text, which have been used as a source of language input for teaching and learning.

3.2.1 The writing of textbooks: general writing and publishing norms

When considering the use of textbooks as a prevailing resource in teaching and learning, it is also relevant to examine *how* textbooks are written and published. Textbooks form a source of content for teaching and learning. Examining their development process(es) can highlight strengths and weaknesses which can be considered for the development of the linguistic content database specification as well as the prototype linguistic content database for the German primary school classroom in Ireland. Points of note include - Do authors and publishing houses use the national curriculum as a basis for their content? In the case of global textbooks, how do they interact with the many curriculums at play in the countries they serve? How relevant is the content to the audience it is produced for? This section takes a brief look at the **how** and **by whom** textbooks are produced globally and in Ireland.

The evidence has shown that textbook authors tend to rely on four different bases when authoring content:

1. their own inspiration and creativity
2. their instinct of what will work well (based on their own experiences)
3. an empirical examination of other resources and their content
4. content development principles or methodologies.

Many textbook writers are teachers with practical experience in the classroom, especially in the language learning domain. They were reported to "rely heavily on their own intuitions, viewing textbook writing in the same way as writing fiction, while at the same time, emphasising the constraints of the syllabus" (Prowse, 2011, p. 158). Many of these writers have been reported to work in an "ad hoc and spontaneous" way relying on an "intuitive feel for activities which are likely to 'work'" (Tomlinson, 2013, p. 97). Furthermore, they have been reported to "rely heavily on retrieval from repertoire, cloning successful publications and spontaneous "inspiration" (Tomlinson, 2012, p. 11). Some of these latter processes involved an examination of "existing coursebooks" and then following on to formulate "a working syllabus" from which to write the textbook content (Popovici and Bolitho, 2003, p. 506). In contrast, other authors have been reported to work within a framework of materials development (e.g. see Jolly and Bolitho, 2011; Tomlinson, 2013) or taken a text-driven approach to materials development (e.g. see Fox, 1998; Tomlinson, 2013; Willis, 2011).

Alongside the use of author *intuition*, authors have also been reported to rely on their theoretical concepts of learning when they write materials (Gottheim, 2010). These concepts can include "concepts of language/foreign language, concepts of learning and of teaching another language" and when combined with intuitive concepts, form an "integrated structure" which are realised through actions and behaviours particular to the teacher (Gottheim, 2010, p. 225). The integrated structure can also be realised through the materials the teacher writes. Variance across teachers' integrated structures would assume much variance across the materials they would all write. Maley (2011) posited further factors which could contribute to Gottheim's notion of teacher variance, including

- "degree of language proficiency and confidence
- previous personal learning experiences
- own personality (introvert/extravert, open/closed etc.)
- preferred teaching style (directive/consultative, etc.)" (p. 379-380).

Teacher textbook authors have reported on the compromises they are required to make when they are provided with direction from the publishing company – where sales potential and market

gap dominate the design and content, where textbooks need to conform to a prescribed length, number of units, content progression rubrics, and where textbooks are required to pass the "flick-test" (Bell and Gower, 2011; Prowse, 2011; Tomlinson, 2008; Mares, 2003). Mares (2003) outlined the struggle he experienced between "the tension that existed between writing materials we believed in and materials someone else believed would sell" (p. 136). He also described the requirement for certain amounts of white space on each page to make coursebooks more user-friendly and visually appealing. Authors are often required to compromise their own beliefs on the strategies and approaches that work well in their own classrooms, as these cannot be transferred to all learning environments where the teaching environment and individual learner are the unknown quantities (Bell and Gower, 2011). In turn, materials drafted for textbooks are often extensively edited when the textbook comes to the editing and design phases of textbook production.

One textbook author described a series of four phases of editing prescribed by the publishing company and editor which was followed by further editing with the design/illustrator team (Prowse, 2011). A further iteration of editing can be required if materials are trialled with teachers in the classroom (Donovan, 1998). The resulting materials may conform to the publisher's, editor's or designer's ideals, but may be far removed from the materials the author had originally envisaged.

In Ireland, the methods employed by textbook authors concurred with findings from abroad. Evidence gathered from Irish publishing companies has reported that textbook authors, who are usually teachers, "draw on their personal experience of what works in practice" (Harbison, 2008, p. 153) when developing textbooks. In some cases, they reverted to what they learned during their teacher training, even if this content or these methodologies were at odds with the current curriculum. They were also reported to draw on their experiences of what they consider "normal practice" and reported to have looked "at textbooks from around the world to get ideas as to how the content of the curriculum could be taught using a variety of approaches" (Harbison, 2008, p. 153).

Harbison (2008) noted that each publishing company's first priority was to sell books, which led to the content, design and production being market-driven rather than curriculum-driven (Bell and Gower, 2011; Harbison, 2008, p. 157; Mares, 2003). Gray (2013) reinforced this point noting that textbooks are "commodities to be bought and sold" and as such, the first priority of publishers is "to earn their producers a living" (Gray, 2013, p. 7 citing Apple, 1985, p. 149). Gray (2013) continued with more details on the market force driving publishing companies:

> "although materials are aimed at use *inside* a classroom, they will always bear the hallmarks of the conditions of their production *outside* the classroom. This is particularly the case with materials which are produced in a commercial context, where the need to maximise sales, satisfy shareholders, and achieve corporate goals may have a direct impact on the design of materials, quite distinct for their pedagogical intent" (p. 7 citing Littlejohn, 2012, p. 284)

This represents somewhat of a paradox, considering that the market for primary school textbooks is created by primary school teachers. Harbison (2008) recorded that each publishing company's second priority was to develop a book which covered the content of the PSC. One of their lesser priorities was a focus on incorporating the content in such a way that it could be taught using the teaching methodologies espoused in the PSC (Harbison, 2008).

If one examines the textbook-dominant classroom and considers the number of agents who interpret the local curriculum, in Ireland, the PSC, before it reaches the children in the classroom, the message of the curriculum could be somewhat skewed by the time it reaches the children, in a manner similar to 'Chinese Whispers'. Harbison (2008) refers to the number of different "interpretations" which the PSC must pass through – (1) publishing company, (2) textbook and (3) teacher – before finally reaching children in the classroom (Harbison, 2008, p. 200). Figure 3.3 below illustrates this phenomenon. Harbison's model does not take account of the pre-publishing 'interpretations' of the author (1) interpreting the curriculum and (2) writing the textbook, as well as introducing the many integrated structures (Gottheim, 2010) which are unique to their personal concepts about language and language learning.

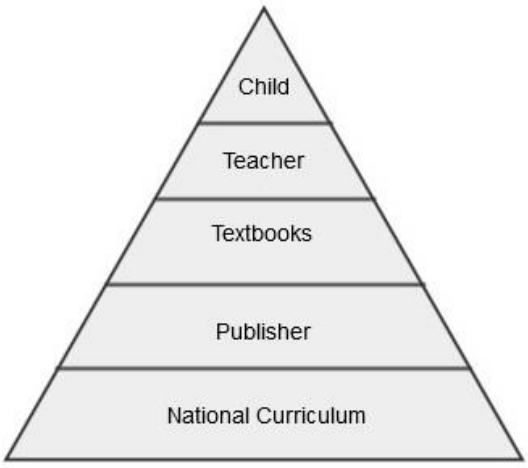


Figure 3.3 Interpretations of the PSC before it reaches the child in the classroom (Harbison, 2008, p. 200)

If one adds the editing dimensions discussed above from the textbook production process, the 'Publisher' level of this triangle already removes the interpretation of materials far from the author's original concept, if this was indeed originally, curriculum-led. In summary, "the intentions of those who write the schemes may not be transferred to the teachers who use them; the way in which schemes are used by individual teachers, as well as the extent of their use, may not reflect these intentions" (Millett & Johnson in Johnson & Millett, 1996, p. 56).

Given the writing and development process outlined above for textbooks, it is no surprise that textbooks can be far removed from the curriculum they are supposed to support. While teacher authors may initially start with the curriculum in mind, the design and publishing processes involved in producing a textbook can often distance the content from its original source. In addition, the authors themselves are presenting their interpretations of the curriculum. These interpretations can be shaped by their own experiences and knowledge and provide an added level of variance to materials supporting the curriculum.

Figure 3.4 below suggests further additions to Harbison's (2008) model such as variance introduced by the textbook author's interpretation of the curriculum, adaptation of content introduced by the publishing processes and the types of variance the teacher's interpretation of the content brings.

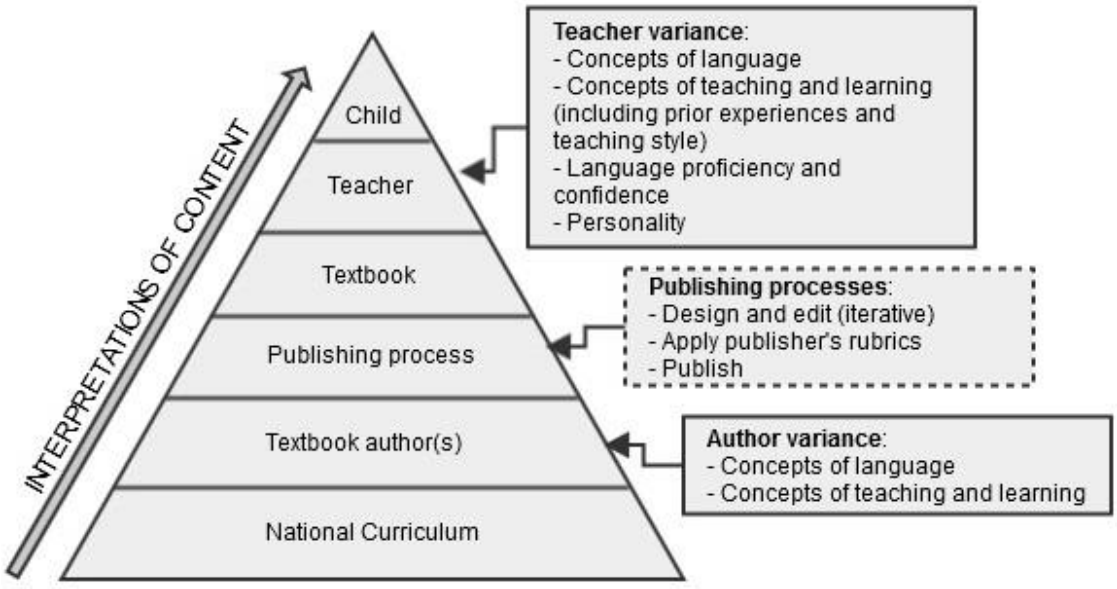


Figure 3.4 Suggested update to Harbison's (2008) model (additional detail on variance informed by Maley (2011) and Gottheim (2010))

Having examined the literature surrounding textbooks from other domains, in part, the acceptance or otherwise of a dependence on textbook use in classrooms is governed by the educational sector or state's viewpoint on whether textbooks form the core of school curriculum

or the curriculum does. Placing textbooks on the levels of localisation spectrum of Figure 2.3 (p. 41) would show their content to be further removed from curricula than originally envisaged, resulting in a requirement for additional localisation for both global and local textbooks. Among textbook writers, many varying personal factors are at work in the development process. Aside from consulting empirical evidence of what other textbooks contain, no other solid development/authoring processes emerge.

The focus now shifts to more formal development models for language learning resources. Many relevant disciplines can inform these processes. The starting point was an examination of disciplines which require **language learning content** or **linguistic resources** as an input or those which specify how these resources should be integrated or produced in some way. "Language" was the required keyword in identifying the relevant educational and material development disciplines. The final selection criteria was to examine domains which focused on producing consumable paper-based resources (e.g. publishing houses or educational practitioners) and consumable digital resources (e.g. Computer-Assisted Language Learning researchers/ practitioners or educational software developers). The (1) traditional language materials development domain and the (2) Computer-Assisted Language Learning domain in themselves, represent specific domains in which language learning content is very relevant.

Within the Language Materials Development domain, language teaching textbooks, coursebooks and software are all outputs of development processes. This domain examines how (methodologies) practitioners and publishers develop, adapt and use language teaching materials. As the domain label suggests, content is to the fore. Considerations such as audience, culture and context are all relevant to the generation of language materials. However, little consideration is ever given to the relevance of prior learning, progression across years, and what can be brought from relevant domains like SLA, or indeed in making the format of the output one which can be reused and tailored to many scenarios. Section 3.2.2 examines development principles and methodologies from this domain.

The Computer-Assisted Language Learning (CALL) domain examines cross-disciplinary research where technology is used to teach and learn languages. With language learning as the basis of the work within CALL, language content (with its origin in appropriate pedagogy) is a requirement for some CALL endeavours, particularly those which develop courseware. Of note, is that the weight placed on the importance of content development within CALL varies significantly from one project to the next. To one extreme, Decoo and Colpaert (1996, 1997) reported how teams at Didascalia/Linguapolis who develop language learning programs, spent "95% of its research and development on content and only 5% on programming" (Delcloque, 1998). Other sources highlighted how content was adapted to suit the technology through which

it was delivered (e.g. content delivered through SMS as discussed in Levy and Kennedy, 2010). Many development methodologies have been proposed in the CALL literature, two of which are discussed in detail in Section 3.2.3 below.

The difficulty with crossing disciplines in search of a suitable framework or methodology is that the focus of the framework or methodology output differs significantly. In the case of Language Materials Development, the output is a form of language learning resource which can be used directly in teaching and learning (a consumable resource, be it paper-based or digital). It is an end product which is delivered to the learner. In this case, the output is rarely in a format which is transferrable or reusable beyond the immediate context or need of the developer (e.g. language learning material is embedded in an illustrated and designed textbook but cannot easily be extracted for use in another format).

Within CALL courseware development – one type of output of using a development model is language learning software/courseware/tools (a secondary output would be learner input to the CALL system). While the principle of "developing and/or using a format for content which is reusable" is a recurring theme across the CALL literature (e.g. Colpaert, 2004; Knapp, 2004; Ward, 2002), the reality can be far-removed from the aspiration. As within the Language Materials Development domain, content can be far embedded in CALL materials. In addition, content is generally only considered at one stage of development where many other considerations are evident in the overall framework or model (e.g. technology to be employed, learner profile and activity type).

3.2.2 Principles and methodologies for content authoring from the language teaching materials domain

Principle-based approaches

It is apparent that similar principles are emerging across disciplines which focus on the teaching and learning of languages: (1) Learning ahead of teaching (e.g. Tomlinson, 2011), (2) pedagogy first (SLA) (Decoo, 1996; Tomlinson, 2011, 2008) and (3) learner to the fore (e.g. Hubbard, 1996; Zilversmit, 1993). These principles echo through the literature on language learning materials development as well as the CALL domain. Within the language learning materials development domain, there is a constant push and pull between these principles and the priorities of the publishing houses which commission and publish the majority of language learning materials.

McGrath (2002) cautioned that principle-based design does not always represent an "objective truth" (p. 155) as even if they are research-based, the principles may not be transferrable. McGrath (2002) reinforced that principles are often derived as a "personal rationale: a key-point justification for the decisions that are to be taken based on beliefs about learning and how this can best be facilitated" (p. 155).

Appendix A provides a listing of some of the principles which emerge from the materials writing literature. The principles can be categorised into themes, which recur across each author's contribution. The themes relate to the visual aesthetic of the materials, the basis of the materials in theories of SLA and learning theory, the characteristics of the materials, the localisation of materials, materials suitability to their target audience and guidelines for authors. Table 3.1 provides additional detail on each of the themes emerging.

The authors associated with some of these principles (some did not specify how the principles had related to their direct development work) outlined that they served the purposes they were employed to do, in developing language learning materials which catered to the specific learning niches they were designed for. No one method proved better or worse than another in that the process was suitable for the author's level of engagement and the output served its destined purpose. Some principles built on the work of others. Tomlinson (2011) provided the most comprehensive set of principles with their basis in SLA. Echoes of his principles can be seen through others outlined (e.g. Farr et al, 2010). However, across the many disciplines which employ language learning materials in some guise or other, many hours are dedicated to developing these materials.

Theme	Detail
Materials' visual/design quality	material impact visual appeal engaging
Materials' basis in SLA and learning theory	use of authentic texts use of authentic tasks explicit grammatical instruction opportunities for learner language use presentation of the four skills progress language learning skills
Materials' characteristics	flexible achieve learner ease
Materials' localisation	cultural linguistic curricular high-tech and low-tech
Materials' suitability to their target audience	teacher student
Authors' writing guidelines	time required assistance from other authors

Table 3.1 Themes which emerge across principles from the materials writing domain (based on principles from Bell and Gower, 2011; Farr, Chambers and O'Riordan 2010; Hann, Timmis and Masuhara 2010; Hutchinson and Waters, 1987 as cited in McGrath, 2002; Nunan, 1988 as cited in McGrath, 2002; Tomlinson, 2011)

The majority of the principles outlined relate to the characteristics that language learning materials (in their end product of textbook or digital resource) should have, and which need to be considered through the course of their development. Only one of these principles is of direct relevance to the development of raw linguistic content, that of ensuring materials are linguistically-localised and localised to the local curriculum (in the case of local materials). The discussion now turns to more specific methodological- and process-based approaches. The focus of the discussion lies in the development of linguistic content, content localisation and reuse of content.

Methodological-based approaches

The methodologies from the materials development domain vary from a series of discrete steps, to a series of processes or methods (e.g. Jolly and Bolitho, 2011; Maley, 2011; Tomlinson, 2013).

Jolly and Bolitho (2011) presented their framework for materials development which focused on the development of an end product for language teaching and learning. However, as it is a process-based framework, many of the processes also transfer to the creation of raw linguistic content. The initial framework proposed was a linear flow-diagram which followed the steps from start to finish. Jolly and Bolitho (2011) added further to the framework to include shortcuts and feedback loops (see Figure 3.5 below), where any one stage of the framework can be skipped or revisited as feedback is gathered and evaluation is carried out. They do this to ensure that

"materials writing can be a dynamic and self-adjusting process" (Jolly and Bolitho, 2011, p 112), through which the materials can be tuned or localised to their users. They stressed that "excellence in materials lies less in the products themselves than in the appropriate and unique tuning for use that teachers might be engaged in" (Jolly and Bolitho, 2011, p.112).

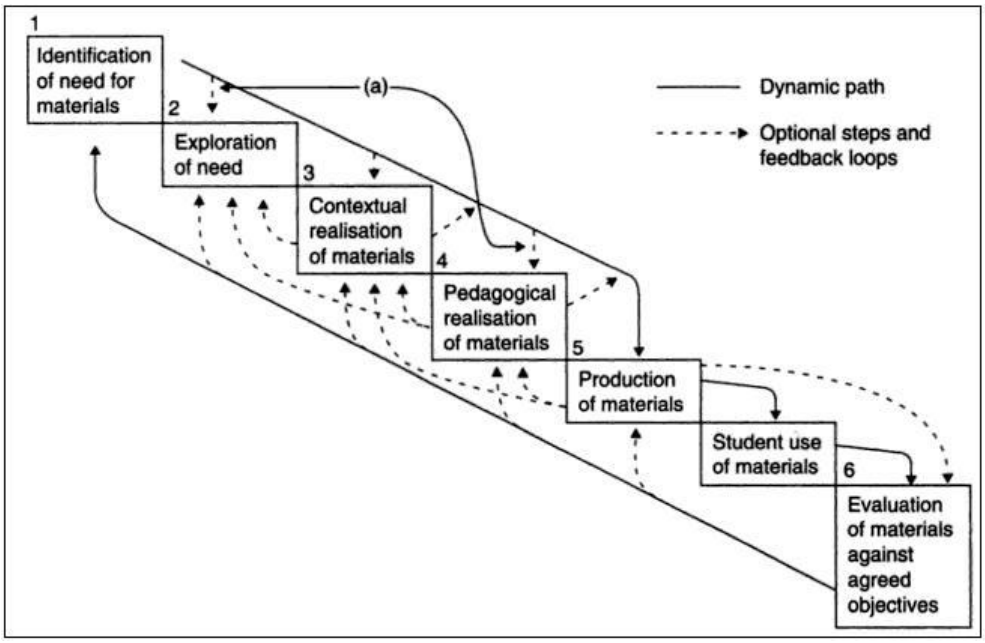


Figure 3.5 "A teacher's path through the production of new or adapted materials" (Jolly and Bolitho, 2011, p. 113)

One final addendum to the framework was found in the appendix where more light was shed on the processes which can underline each of the stages of the framework. Jolly and Bolitho (2011) noted that "Questionnaires; feedback from students in class; formal or informal diagnosis of errors and shortcomings in learners' competence; analysis of existing course materials; pre-course needs analysis" (p. 131) can all feed in to identifying the initial requirement for materials as part of the development framework. Jolly and Bolitho (2011) accounted for reuse of materials in their framework through the adaptation of materials. While the framework intent of this adaptation is the localisation process, it also allows for materials reuse in this way.

Tomlinson (2013) presented two text-driven frameworks for materials development (the latter one is an adaptation of the first to account for web-based texts). They are practical frameworks which delve more into the associated activities which can be explored around any given text to help learners to learn through the text, rather than a higher level- or meta-view of materials development. Generating activities around a selected text fills the criteria of localisation of materials and reuse of materials. Associated activities are developed around the text which has been selected but also reflect the personal preferences and needs of the learners as well as drawing on their own life experience to recall emotions or experiences through which they can

identify with the actors in the target text (e.g. readiness activities, experiential activities and development activities). Table 3.2 below summarises the stages involved in Tomlinson's (2013) framework. The reuse of materials is apparent through the use of a pre-existing text and the same series of activities through which learners work with each new text.

Framework Stage	**Description**
Text collection	Gather texts which you feel will meet a learning need. These can be built up into a library of texts for later selection.
Text selection	Choose a text from your library which will be engaging for learners but also conform to selection criteria listed by Tomlinson (e.g. "Are the target learners likely to be able to connect the text to their lives? Is the linguistic level of the text likely to present an achievable challenge to the target learners?" (Tomlinson, 2013, p. 101)).
Text Experience	Re-read and re-examine the text in order to re-engage with it. Draw on learner's personal experience to identify with elements of the text.
Develop Readiness Activities	Learners work with activities which will help them to engage with the text they are about to read/listen to.
Experiential Activities	Learners place themselves in the text being explored as the main agent/subject or internalise the text by associating with the emotions or lessons encountered in the text. Alternatively, learners can contribute to the construction of text.
Intake Response Activities	Learners are asked to reflect and focus on what has been taken in from their first encounter with the text.
Development Activities	Learners work with "activities which provide opportunities for meaningful language production based on the learners' representation of the text" (Tomlinson, 2013, p. 105 citing Tomlinson, 1999).
Input Response Activities	Learners revisit the text and examine its "purposes" and the "language" within it (Tomlinson, 2013, p. 105).
Trialling	These final three stages are added to the framework summary provided by Tomlinson to reflect the practitioner using, evaluating and revising chosen texts and associated activities.
1. Evaluation	
2. Revision	

Table 3.2 Summary of Tomlinson's text-driven approach to materials development (2013)

The second of Tomlinson's frameworks (2013) is a web-based adaptation of the first. Texts in this adapted instance can be libraries of texts online, compiled by the practitioner/teacher or learner. Alternatively, they can be sourced online by the learner to meet the criteria set out by the teacher. The same principles of localisation and reuse apply to the web-based framework as to the initial framework discussed above. The missing component of the research review is raw linguistic creation. As both frameworks draw on existing authentic texts, there are no processes considered for (raw linguistic) content creation (rather selection).

Maley (2011) was discussed in Section 2.5.1 in the context of the localisation of consumable paper-based resources. He suggested two ways to ensure materials are localised for their learners. The first was ways through which existing materials could be adapted to "bridge the [localisation] gap" (p. 380) between learning materials and the learner. These processes were ways through which the teacher could localise the materials for his/her leaners (e.g. reorder, rewrite/modify, omission or reduction). The second of these was a new way to design and write materials which provided enough choice for the teacher to be able to select those which are most appropriate for their setting, thus fulfilling the research criteria of reuse. The content created through this process is pre-formed "written, spoken or visual texts" (p. 385), which do not correspond to the format of raw linguistic content. Maley's (2011) second method is based on proposals made by Prabhu (1988), and are discussed below.

Maley (2011) provided the rationale for a new way to design materials which provide greater control to the teacher in some of the "four major factors in the classroom: Content (what), Order (when), Pace (how fast) and Procedure (how)" (p. 385).

Maley (2011) discussed Prabhu's suggestion for coursebooks which comprise semi-materials and meta-materials. **Semi-materials** are essentially the content input, and can be

- single-type: activities and exercises which can be chosen in any order for use by the teacher (e.g. writing activities and vocabulary development work)
- collections of raw input: collections of "written, spoken or visual texts" (p. 385) which are available for use without any instruction on how they should be used (e.g. vocabulary in context, role-play and comprehension questions).

Meta materials are essentially the procedure for use with the input, and were described as "empty pedagogical procedures" (e.g. dictation, role-play and gap-fill) (Maley, 2011, p. 385). These can be matched with semi materials to form a teaching exercise/unit.

Maley (2011) extended Prabhu's description to allow for "**flexi-materials**" which provides a listing of semi- and meta-materials from which the teacher can pick and choose combinations suitable for their learners. This system allows the teacher to decide the order, pace, procedure and choice of text in each case (Maley, 2011). Maley stated that the flexibility lies in the ability to reuse materials and procedures time and time again, as one text type could be reused through many different procedures for instance. Maley (2011) cited the potential computers have to play in this process with the addition of multimedia-based materials adding to the possible permutations of materials as well as providing quick and easy access to the materials. Use of computers would allow the materials to extend beyond the boundaries of the coursebook.

Piloting of materials (formative evaluation) and feedback loops

Having seen how different approaches have been developed for language materials generation, it is also important to recognise that within many of the domains of language materials development, there are piloting stages which test out materials in the classroom. If it is the teacher/practitioner developing their own materials, then the materials are used with their own students and feedback is incorporated into any revision to materials (Tomlinson, 2013).

In the case of publishing houses, a piloting stage is generally incorporated into the development of new materials (Donovan, 1998). Donovan (1998) described how most, if not all materials which are being drafted for publication will be piloted by the author with their own students, by associates of the author, or by teachers in schools who sign up to trial materials. Donovan (1998) also described how the extent of the piloting phases depended on the volume of materials being piloted, whether there is more than one book of materials in the series to be piloted (e.g. coursebook, plus workbook, plus teachers' manual) and the budget or time constraints which may need to be considered. Mares (2003) similarly described how time and budget constraints can delimit the amount of time which can be attributed to piloting materials, as can the privacy which sometimes needs to be maintained around the piloting of new materials to ensure that any novel ideas are not copied by other companies. Difficulty in attaining willing teachers and practitioners to pilot materials was cited by both Donovan (1998) and Mares (2003). The difficulty of accessing busy teachers and practitioners for any piloting or review work should not be underestimated. Questionnaires were generally used to gather teachers' responses to the piloting phase (Donovan, 1998; Mares, 2003) and ensure materials were fit for purpose.

Popovici and Bolitho (2003) described the first time that language learning materials were piloted in Romania. Traditionally, textbooks had been "published … without prior consultation with students and teachers" (Popovici and Bolitho, 2003, p. 508). Popovici and Bolitho (2003) described the large scale piloting that targeted practitioners in all "40 counties of the country" (p. 508). As other editions in the series of the textbooks were being produced, the piloting became less extensive albeit more focused. In congruence with Mares (2003) and Donovan (1998), Popovici and Bolitho (2003) cited "organizational, financial and competition-related reasons" (p. 509) for the reduction in the scale of the materials piloting. They did however note that "more targeted questions" and additional expertise amongst the development team in analysing feedback, made up for the reduced scale of the pilot sample (Popovici and Bolitho, 2003, p. 509).

In summary, Tomlinson (2013a) compiled the processes outlined by authors who had developed language materials for teaching specific skills and all skills. There is some overlap in the processes outlined, as needs analysis, for example, may already identify the required syllabus. Similarly, some of the processes which are embedded within the listing, present some overlap

in the process (rather than the material being produced). An example of this is "trialling sample units" and "trialling the materials" (Tomlinson, 2013a, p. 537-8).

a. "Needs analysis of the target users
b. Determination of pedagogical approach
c. Determination of frameworks for developing the materials
d. Determination of the syllabus
e. Drafting sample units
f. Trialling the sample units
g. Revising the syllabus, approach and frameworks
h. Finding and/or developing texts
i. Producing the materials
j. Monitoring the materials
k. Trialling the materials
l. Revising the materials
m. Editing the materials."

Tomlinson (2013a, p. 537-8)

The cumulative processes start to resemble a more structured development methodology rather than the looser principle-based means or framework based-methods discussed. Processes (a) and (d) can be related to localisation, (e) and (i) to the creation of content (with (j) to (m) part of their development process) and (h) relating to the reuse of existing texts. Although these processes have not all related to the exact match of what is required in this research instance, they assist in informing the processes which can underlie their creation and integration. The next sections discuss formal development methodologies from the Computer-Assisted Language Learning domain with the same three focuses of content creation, reusability of content and localisation of content.

3.2.3 CALL courseware development methodologies

Computer-assisted Language Learning (CALL) is a multi-disciplinary field which draws from many fields of inquiry. As such, the design and development of CALL courseware draws from these disciplines and includes Human-Computer Interaction (HCI) (Hémard, 1997; Peterson, 1998), Second-Language Acquisition (SLA) (Chapelle, 2001) and software engineering (Colpaert, 2004). HCI reflects the user-centred design central to any CALL development process. Many methodologies for a specific CALL courseware design have been proposed in the literature (Colpaert, 2004; Decco, 1984; Hubbard, 1992; 1996; Meskill, 1999; Underwood, 1984; Watts, 1997), all integrating different methodologies from related disciplines.

Two particularly important and widely adopted CALL methodologies are Hubbard's Elements of CALL Methodology (1996) and Colpaert's Research-Based Research-Orientated (RBRO) Model (2004). Despite being nearly 20 and 10 years old respectively, they have both prevailed across the literature as courseware development methodologies. This could be the case as the scope for digital tools which could be used to support language learning (e.g. CMC, social media, mobile learning, web-based resources, corpus-based resources, apps) exploded around the year 2000, with a resulting reduction in focus on complete courseware packages. The focus of CALL research shifted to the development of smaller language learning packets or the impact of the use of these newer digital facilities on teaching and learning. Both development models remain relevant as they map the development path for CALL courseware development. Neither is too specific to be outdated due to new affordances offered by technology today.

Hubbard's framework (1996) offers a series of considerations for courseware development within three modules – Design, Evaluation and Implementation. Each of the three modules interact with and inform one another. Being a framework rather than a development model, it is written in more generalised terms than Colpaert's (2004) RBRO-Model, and guides the developer through areas of consideration in CALL development. Hubbard (1996) merits that his methodological framework provides a map, which helps the learner, teacher or software developer to disambiguate the CALL development process, so that they can identify CALL solutions to their own learning, teaching or development problem or need.

Colpaert's RBRO model on the other hand, is a development methodology. It offers a more structured and definitive path from start to finish in the development of CALL courseware through the stages of Analysis, Design, Development, Implementation and Evaluation (ADDIE). As a waterfall model, the output of each module feeds in as the input to the module following it. The difference between the two models can be ascribed to the differing definitions of what constitutes a framework and a methodology. A framework is a scaffold or outline which guides the processes which need to be conducted to reach a certain output. A framework provides parameters or guiding principles rather than a set of specific steps or processes which must be followed. A methodology is a more specific and definite entity which outlines the exact processes which need to be followed to reach a certain goal or develop a particular output. A methodology will ensure that many individuals can reach the same goal by following a very specific route.

Both Hubbard (1996) and Colpaert (2004) place great emphasis on the human end of CALL courseware development. Hubbard highlights this in the introduction to his framework – "often missed is the fact that the field really involves the interplay of humans and technology and that the human end is especially significant" (1996, p. 15). The teacher/practitioner's didactic choices

and the learner profile are all considered in the Development Module. Hubbard also ensures the human element is to the fore in the Evaluation Module where the "determination of fit" is partly deemed according to the "learner fit" and "teacher fit" (and finally, "operational description") (1996, p. 27). As the inverse of the Development Module, the classification of 2/3 of the development criteria into these categories in the Evaluation Module proves the significance of the human role in Hubbard's model. In congruence with this take on CALL courseware development, Colpaert (2004) considers the learner and teacher within the GLDT grid (General, Local, Differential and Targeted requirements) of the Analysis Phase and uses personas to realise solutions to identified needs in his Design Phase. The human priority is a significant parallel with the work outlined in this thesis. In order to localise teaching and learning resources and content to their learners, the learner and his/her immediate needs and context are the most significant factors for consideration.

Both methodologies were realised through the compilation and consideration of other relevant methodologies (Hubbard, 1996; Colpaert, 2004) and in Colpaert's case (2004) on "existing concepts" (p.104) in language learning rather than newly developed ones. This research attempts a similar process of building a solution to identified challenges by examining and reusing relevant methods from other practitioners and researchers. While there are similarities and differences across Hubbard's (1996) and Colpaert's (2004) models, the purpose of the following discussion is not to analyse the models for incompleteness, rather to extract relevant areas for consideration in the design and development of the linguistic content database. While both methods involve content creation for language learning, neither one is exclusively designed for this process. The coming sections examine the processes involved in each method and point to the relevant areas for extraction and consideration in the next chapter concerning research context, design and methodologies.

Hubbard: Elements of CALL Methodology (1996)

Hubbard (1996) has developed a framework which aims to capture the main "methodological issues surrounding the use of software that includes content designed or adopted for language learning purposes" (Hubbard, 1996, p. 15) or the design of *courseware.* Hubbard based his methodology on the previous methodological descriptions of Richards and Rodgers (1982) whose focus was on language teaching defined according to approach, design and procedure and Phillips (1985) whose work used seven categories to form a "typological framework for describing the form, content and implementation of CALL materials" (Hubbard, 1996, p. 18). Hubbard's methodology highlights the importance of taking technical considerations along with language teaching and learning considerations into account and stresses that the "human end [of CALL] is especially significant" (Hubbard, 1996, p. 15).

Hubbard (1996) outlines the four main *players* who contribute to the human element of his methodology: (1) the learner, (2) the developer, (3) the evaluator and (4) the classroom teacher. He stresses the importance of removing the technological focus of CALL courseware and bringing the locus back to a more human one through consideration of the roles of the four players in courseware development. Hubbard's methodology involves three mutually impacting modules - the development, evaluation and implementation modules (Figure 3.6).

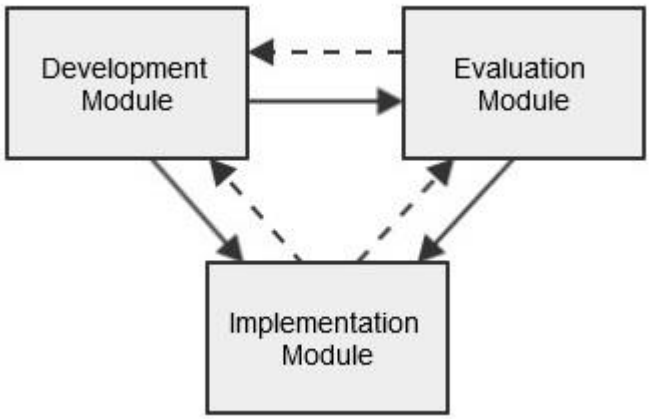


Figure 3.6 CALL Methodological Framework (Hubbard, 1996)

Hubbard divides the **development module** into *approach, design* and *procedure* (Figure 3.7). The *approach* is informed by the teacher/practitioner and includes items like language teaching approach (for example, communicative or audio-lingual approach). The *design* is informed by both the teacher/practitioner and the learner and includes elements such as learner profile, syllabus and programming languages. The *procedure is* technology-related and involves aspects such as screen layout, feedback and activity type. Feedback refers to the type of feedback the CALL courseware would provide.

Hubbard's methodology (1996) recognises the importance of content development within the design phase of the development module. This phase of work represents the greatest area of fit to this research. Hubbard emphasises how elements such as learner profile, syllabus, language difficulty, (language) program difficulty and content all feed into materials development for courseware production (see red highlight in Figure 3.7 below). He considers 'content' separately from the other design elements, in the same way that raw linguistic resources are also abstracted away from their surrounding product. The syllabus, language difficulty and program difficulty all interact with the content to inform its development. These elements are all relevant to defining the context of a learning environment so that relevant localisation can be considered.

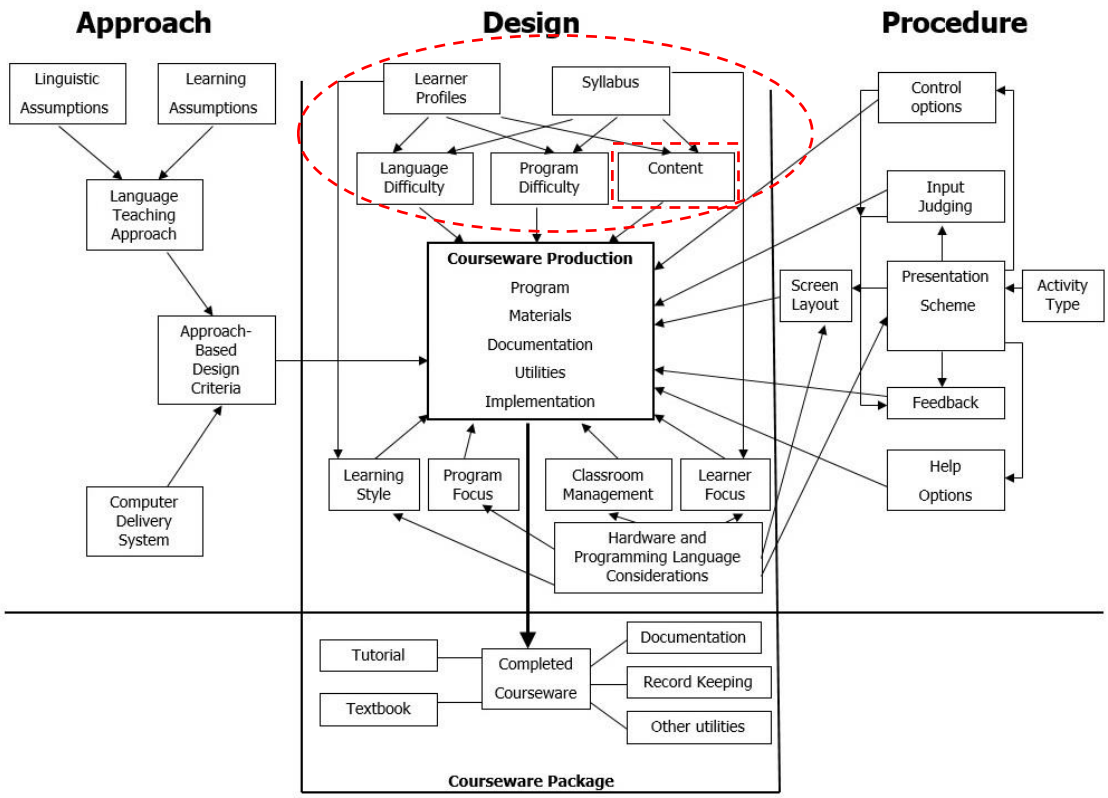


Figure 3.7 Hubbard's Development Module (Hubbard, 1996, p. 19) (red emphasis added to highlight relevant areas for constructing raw linguistic content and considerations for localisation)

Elements for consideration within the "learner profile" include the "learners' proficiency level, age, native language, needs and interests" (Hubbard, 1996, p. 22). Other relevant considerations from Hubbard's design phase include those factors which he describes as determining the "language difficulty" – familiarity, concreteness, length, grammatical complexity and clarity of signal (Hubbard, 1996, p. 22). Elements of consideration in Hubbard's model such as the language teaching and learning style of the learner are exclusive of the development of raw linguistic content, as it is the teaching practitioner who brings these factors into consideration.

The **evaluation module** is viewed as the inverse of the development module. The development module defines criteria (some of which were outlined above). The system is then developed according to these criteria. The evaluation module on the other hand, looks at the developed courseware and evaluates whether all of the criteria set out in the development module have been fulfilled in the courseware. The evaluation module contains elements which correspond to those in the developmental module. Procedure is referred to as *operational description*, design as *learner fit* and approach as *teacher fit*. If these three elements are evaluated positively, then an *appropriateness judgement* is obtained. Further implementation can proceed until the final product is completed.

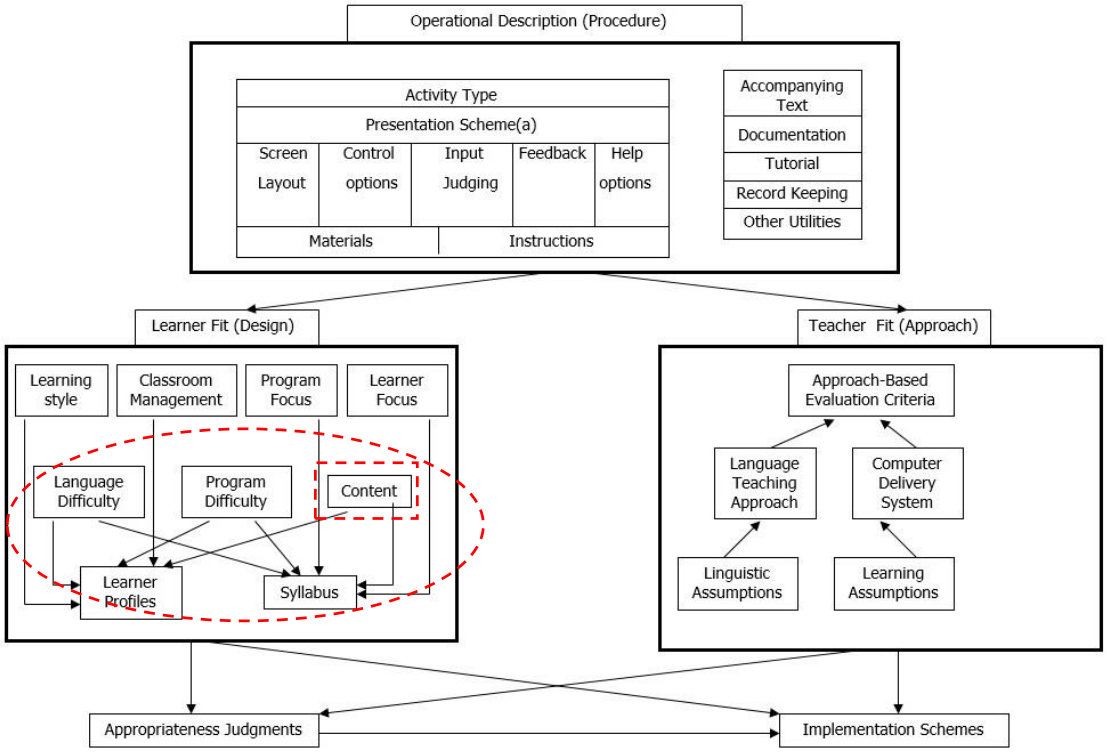


Figure 3.8 Hubbard's Evaluation Module (Hubbard, 1996, p. 26) (red emphasis added to highlight relevant areas for evaluating raw linguistic content and considerations for localisation)[8]

Content remains a discrete element in Hubbard's evaluation module within the criteria for 'learner fit' rather than being consumed under the 'materials' banner of the development of courseware. This again emphasises how content is a separate entity and consideration in evaluation. The same elements of the 'learner fit' are relevant to the localisation of content – language difficulty, program difficulty, content, learner profiles and syllabus (see red emphasis in Figure 3.8). The remaining elements making up the 'learner fit' are more relevant to elements which the practitioner decides and brings to the educational setting.

As with the previous two modules, the **implementation module** is divided into subheadings, which include "the flow of the CALL lesson" and "teacher control". The flow of the CALL lesson includes preparatory activities, tutoring and content preparation. Preparatory activities involve what users must do in familiarising themselves with the courseware before they can sit down and use it. The teacher control involves those elements of a CALL program that the teacher is able to adapt and have an impact on.

Within Hubbard's Implementation Module, emphasis is placed on accompanying preparatory materials and follow-up activities before and after a courseware 'lesson.' He stresses how a

[8] Hubbard (2006) presented some updates to the labelling of the elements of the Evaluation Module.

teacher most commonly has to produce these materials outside of the courseware environment, and makes a case for their inclusion within the courseware. In this instance, the use of raw linguistic content could be relevant as the same raw linguistic content could be presented in different ways in preparatory and follow-up activities – the content remains the same to foster recycling and revision, but the output format differs. The importance of "supporting review, reflection or synthesis" (Hubbard, 1996, p. 31) within the learning process is recognised by Hubbard. Table 3.3 below summarises the areas of relevance and crossover between Hubbard's methodology (1996) and the three resource challenges – abstraction of or discrete consideration of raw linguistic content, reusability of content and localisation of content.

Elements of CALL Methodology (Hubbard (1996))		**Separate content development**	**Reusability of content**	**Localisation**
Development Module	Design Phase: Learner profiles, syllabus, program difficulty, content	YES	N/A	YES
	Design Phase: Language difficulty - familiarity, concreteness, length and grammatical complexity	YES	N/A	YES
Evaluation Module	Learner fit: program focus, learner focus, language difficulty, program difficulty, content, learner profiles, syllabus Teacher fit: Linguistic assumptions	YES	N/A	YES
Implementation Module	Accompanying preparatory materials Accompanying authored follow-up materials	N/A	Reuse of raw linguistic content in alternative formats	N/A

Table 3.3 Examining the three resource challenges in Hubbard's methodology (1996)

Colpaert: RBRO-Model (2004)

Colpaert (2004) provided an extensive review and meta-analysis of development methodologies for CALL courseware up to 2004 and described his RBRO-model for courseware development. In formalising his RBRO-model, Colpaert draws in features of software engineering to produce a CALL development model which still prevails, ten years after it was first published (e.g. Greene, 2013; Ward, 2013). Content considerations weave through the RBRO-Model description and surrounding discussion in Colpaert's work (2004)

The RBRO-model stands for Research-Based Research-Orientated model. It is structured around the well-established ADDIE model from the software engineering domain, which is a waterfall model where each stage of the model – Analysis, Design, Development, Implementation and Evaluation – flows into the next (Sommerville, 2004). Each stage of the Colpaert's RBRO-

model (see Figure 3.9 below) produces an output which in turn is the input for the next stage. "Design is considered to be a working hypothesis which can be verified and validated (or adjusted) after each implementation and evaluation. Each iteration leads to a new working hypotheses" (Colpaert, 2004, p. 135).

In following an ADDIE model, the framework for Colpaert's RBRO-model follows the stages of Analysis, Design, Development, Implementation and Evaluation. Colpaert (2004) refers to each stage in the model as a module.

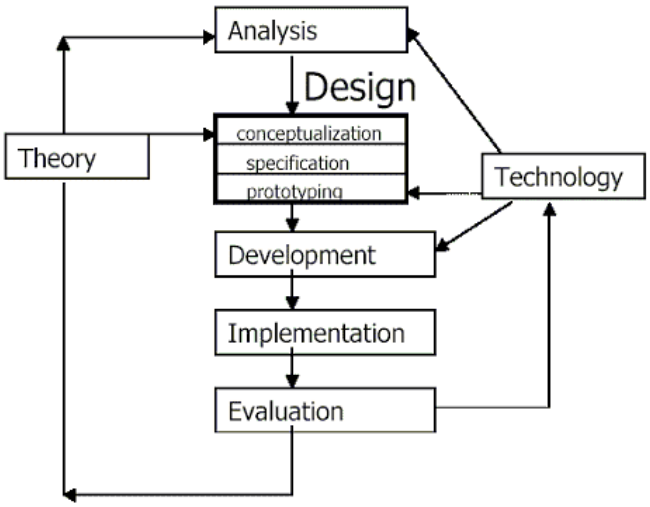


Figure 3.9 The RBRO-Model (Colpaert, 2004, p. 135)

The **Analysis** module is considered the pre-design phase. Its output supplies the input to the design module. Information is gathered during this phase relating to "all possibly relevant epistemological, empirical, actorial, contextual, technological, feasibility-related and perceptive aspects, findings, principles and considerations" (Colpaert, 2004, p. 133). These areas represent the core factors which had been analysed in previous CALL development projects Colpaert reviewed in the CALL literature. He divides this information into three domains:

- *Interdisciplinary expertise:* this is the knowledge developers should have about language, language learning, CALL, technology and architectures, software engineering, HCI etc.
- *Knowledge of the design space:* developers must form accurate ideas about the learner, teacher, and other actors, the available infrastructure, pedagogical traditions, language method, textbooks, language learning situation and the sociocultural background.
- *Identification of the target:* what is the system supposed to change? What can be changed?

Background knowledge relevant to the design space is explored through the analysis of *Interdisciplinary Expertise* and *Knowledge of the Design Space* which are both relevant to the localisation of content and resources. Once all three of these factors have been described, they

are converted into *an operational grid of system requirements.* The grid is based on 4 levels of requirements (GLDT):

1. "**General requirements** include both high-level and lower-level considerations that must be considered in relation to language courseware engineering in general.
2. **Local requirements** are circumstances which are specific to a particular design space or context, and which must be respected as such
3. **Differential requirements** are parameters, which should allow the system to adapt to several possibilities within a particular context or design space, or to changing circumstances…
4. **Targeted requirements** are factors amenable to improvement, which means the aspects that can and should be improved by the system that will be developed…"

(Colpaert, 2004, p. 138) (see Table 3.4 for details).

The local requirements examine those elements which are relevant to localisation within this research. Similarly, the differential ones are relevant to the reuse of content in alternative settings ("which elements are subject to change" (Colpaert, 2004, p. 138)). Within the GLDT grid Colpaert places specific emphasis on the *learner* and *content* in the instantiation of the GLDT grid (among four other considerations – teacher, pedagogy, technology and other actors) – which are relevant to CALL courseware development, but not to raw linguistic content development as it is exclusive of pedagogy (which is chosen by the teacher on application of raw content). The former of these, the learner, is a very relevant consideration to the localisation of content. The latter, content, is a transferrable consideration for raw content development. Content is considered separately from the other considerations, as raw linguistic content is similarly abstracted away from its surrounding product.

	General	Local	Differential	Targeted
Learner	What are generally accepted findings and principles for language learners?	What are common characteristics of the language learners in this particular design space?	What distinctions must be made within this design space or which elements are subject to change?	Which characteristics are amenable to improvement (vocabulary, topics, skills, etc.)?
Teacher	What findings and principles are generally accepted for language teachers?	What are common characteristics of language teachers in the design space?	Which distinctions must be made within this design space, or which elements are subject to change?	Which characteristics are amenable to improvement?
Pedagogy	What are generally accepted findings and principles for language learning and teaching?	Which language learning/teaching method is currently being used within the design space?	Which distinctions must be made within this design space, or which elements are subject to change?	Which aspects are amenable to improvement?
Technology	What does a SWOT analysis of technology in language learning show? What does a SWOT evaluation determine about existing comparable language courseware?	What infrastructure and equipment are available in the design space? What language courseware has been used before?	Which distinctions have to be made within the design space, or what is subject to change (e.g. operating system, network types, processor type, software versions, etc.)?	Which aspects are amenable to improvement (e.g. less network traffic, faster execution, etc.)?
Content	What content is available worldwide?	What kind of content is being used in this design space (textbook, syllabus, etc.)?	Which distinctions must be made within this design space, or which elements are subject to change?	Which aspects are amenable to improvement?
Other actors*	What is their generally accepted overall role during implementation, use, and evaluation?	What is their presence and role in the design space?	Which types of actor can be distinguished and which actor characteristics are subject to change?	What is amenable to improvement (teacher guidance, parent control, mediated communication with native speakers, etc.)?

* Other actors: content providers, native speakers, parents, training managers, software providers, policy makers…

Table 3.4 Explanatory GLDT (General, Local, Differential, Targeted) Grid (compiled from Colpaert, 2004)

Input to the **Design** module comes from the output of the Analysis module. This module is divided up into three stages:

- *Conceptualisation*: the elaboration of a concept
- *Specification*: description of the system's architecture in terms of components and their interaction
- *Prototyping*: testing of discrete topics using available technologies

Conceptualisation is the process of translating system requirements to solution concepts. This is achieved in the RBRO-Model by focusing on pedagogical factors rather than technological ones. In Colpaert's view, conceptualisation consists of "two concurrent and iterative activities: concept development and the application of usefulness criteria" (Colpaert, 2004, p. 140) (see Figure 3.10). He stated that a 'metaphor' is often useful in allowing a synthesis "of a large number of interrelated actors and factors…Examples of metaphors are: an interactive textbook, an adventure game (treasure hunt), or a virtual campus" (Colpaert, 2004, p. 143).

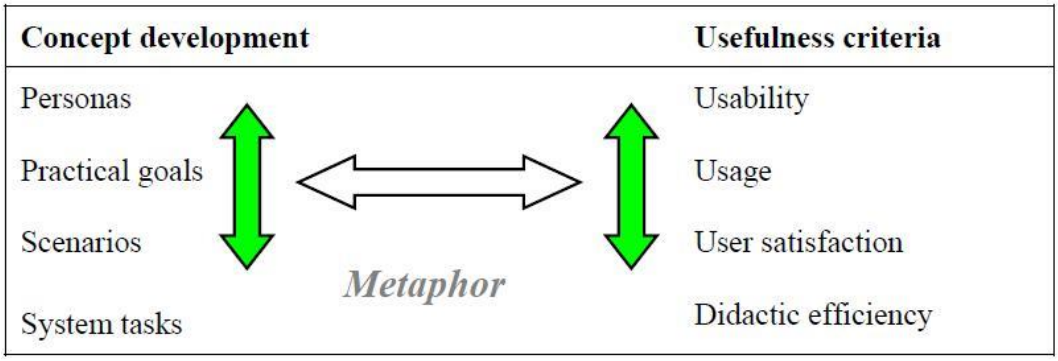


Figure 3.10 Conceptualisation (Colpaert, 2004, p. 140)

Colpaert discussed the Conceptualisation process in great detail, and proposed a "generic and comprehensive construct" (*GCC)* (Colpaert, 2004, p. 149) to realise a technological solution to the analysis of a design space. Colpaert used an analysis of 20 language learning programs to itemise possible technological solutions into a grid comprising "four levels (application, modus, document and item), three interaction types (user input, system output and system processing) and a varying number of user types (learner, designer, parent, etc.)" (Colpaert, 2004, p. 185). This framework/itemisation formed the basis of the GCC where the correct solution can be mapped from the identified needs in the design space.

Specification details the back-end and front-end of the system - the back-end being the system structure in terms of components and their interaction and the front-end representing the user interface with screen design, menu systems and navigation. Colpaert specifies three

tiers for describing the specification: (A) User interface layer, (b) Business Layer and (C) Data Layer. Tier A is self-explanatory. Tier B comprises the interactions between each of the users of the system with the system. Tier C is the application database and provides the content for the system. The concept of a didacteme is introduced within the bounds of Tier C, which has been defined as "the minimal learning unit on which a learner must concentrate for mastery" (Decoo & Colpaert, 1997, p. 68). This specification of Tier C is relevant beyond the three resource challenges being identified, and to the development of the database specification in this research. Maintaining a database of content rather than embedding it across the end product allows for the easier reuse of content.

The final stage of Design is *Prototyping*. This stage involves the development of 'discrete' elements to test their feasibility (with a new technology for example).

Colpaert (2004) places the 'persona' to the fore in the design process, that is, "pretend users…Personas are not real people, but they represent them throughout the design process. They are hypothetical archetypes of actual users" (Colpaert,, 2004, p. 126 citing Cooper, 1997, p. 123-4). Colpaert emphasises the importance of personas in linking the "system requirements and the concept of the solution" (Colpaert, 2004, p. 126).

Colpaert (2004) points out that the Design module is labour intensive but that substantial savings can be made in the Development module if enough effort is invested at the Design stage.

The **Development** module is the 'execution' of the Design plan. The Design is coded. This stage also involves testing. Colpaert (2004) outlines the testing used at the Didascalia/Linguapolis research centre:

Pre-testing: specification checking. Performed by the programmer. Feasibility is judged and any foreseeable problems between linguistics, didactic and technical specifications or requirements are identified.

- Testing 1: Routine Checking. Programmer double checks each routine himself.
- Testing 2: Content implementation testing. Content developers check if program is functioning correctly at each step in content development.
- Testing 3: Operational testing/debugging. Paid external users test the program and try to create conditions that might make the program crash.
- Testing 4: Content testing. Native speakers make corrections and give suggestions.
- Testing 5: Beta-testing with real users in a real environment. Results of the testing phase lead to a new finalised version of the program.
- Testing 6: Real-world testing. Product is tested and launched on the market.

- Testing 7: Research evaluation. Many researchers evaluate existing related courseware as part of their research.

Testing 2 (Content Implementation Testing) and Testing 4 (Content Testing) are relevant to the development of raw linguistic content. The content is retained as a separate entity during these testing phases. Further testing of content in Testing 5 (Beta-testing with real users) signifies an important step in the creation of content – testing it with real users in a localised context. Colpaert embeds the content in the CALL courseware under testing at this stage, but it can still be accessed and extracted for reuse.

The **Implementation** module refers to the use of the courseware in the environment it was designed to cater for. Colpaert (2004) argues that this stage should be successful if Analysis, Design and Development stages were completed comprehensively.

Colpaert's **Evaluation** module is a summative form of evaluation which is carried out to "formulate a new working hypothesis for the next developments" (Colpaert, 2004, p. 146). Formative evaluation is replaced in this model through iterative prototyping, ensuring the product is fit for purpose. Results of summative evaluation feed into future developments in an engineering loop, rather than the current one.

Within the delimits of Colpaert's RBRO-Model description (Colpaert, 2004, p. 135-146) and the surrounding extension of the description across his thesis, relevant content considerations weave through the discussion. Table 3.5 below summarises the main areas of relevance from Colpaert's RBRO-Model to the development of raw linguistic content, the localisation of content and the ability to reuse content.

RBRO-Model (Colpaert (2004))		**Reusability of content**	**Separate content development**	**Localisation**
Analysis Module	Analysis: interdisciplinary expertise, knowledge of the design space and identification of the target	N/A	YES	YES
	Analysis: Local Requirements	N/A	YES	YES
	Analysis: Differential Requirements	N/A	YES	YES
Design Module	Personas in the design process to realise solutions to analysed needs	N/A	YES	YES
	Specification: Tier C	Provides basis for considerations of database representation, storage and usage	YES	N/A
Development Module	Testing 2, 4	YES: Database-based	Formative evaluation	N/A
	Testing 5	YES: Database-based	Formative evaluation: Raw linguistic content + localisation embedded in learning activity or resource	
Implementation Module	Efficient Implementation realised from the Analysis and Design Phases	Realisation of localised raw linguistic content database which maintains reusability of content		
Evaluation Module		YES: Database-based	Summative evaluation	

Table 3.5 Examining the three resource challenges in Colpaert's RBRO-Model (2004)

Colpaert (2004) also reviewed those elements from other CALL development methodologies which need to be considered in content development. He outlined four demands which content must meet in order to be suitable for language courseware development (Colpaert, 2004, p. 59-63):

1. Content quantity: Colpaert outlined how content volume for language courseware is often underestimated.

2. Content quality: He summarised the literature to outline the main characteristics that content should have – it should be rich, varied, correct, authentic, accurate and engaging, complete and comprehensible.
3. Content development: Colpaert recognised that content development is a labour intensive process which requires (1) content authoring, (2) content structuring, (3) content formatting, (4) content input, (5) multicarrier output, (6) content testing and debugging (to include native, didactic, user and software checks) and (7) content updating.
4. Reusability of content: Colpaert maintained that content embedded in courseware should be easily extractable. He emphasised how "a great contribution to CALL research would be a generic database structure that would guarantee the reusability of content"
   (Colpaert, 2004, p. 63).

Hubbard's (1996) and Colpaert's (2004) models offered development insights into the creation and testing of raw linguistic content and the considerations which need to be borne in mind to ensure that it is localised to a particular context. Both models maintain and consider content separately in the development and evaluation processes, with content remaining a discrete and abstracted database across Colpaert's model from its specification in the design phase. Neither model provides much insight on the very elemental processes of authoring content from scratch. However, insights into these processes were gleaned from the discussion surrounding the principles and methodologies from the materials development domain. One further approach to content authoring or development is discussed in the next section – the employment of corpus-based linguistic resources.

3.2.4 Data-driven approaches to materials development: use of corpora in language teaching and learning

Corpora are electronic collections of texts. They can vary in size from 100 million words as in the British National Corpus (BNC) to the 650 million words as in the Bank of English. Corpora provide a huge repository of language in use and in context, and can be used directly to assist language learners to find examples of vocabulary in context, to examine grammatical features, or to examine collocations (words which naturally occur together more often than statistically 'normal').

By their very nature, corpora already represent content databases for language learning, although not generally originally compiled to be applied to a direct language learning task. Their content requires some processing to extract language learning content for application

to language learning. The use of lexical analysis software or corpus retrieval software (such as WordSmith tools, for example) facilitates direct access to the contents, generally, for more advanced language learners.

McEnery and Xiao (2010) cite Leech's (2005) trichotomy of corpora in use for language learning and teaching –

> "the indirect use of corpora in teaching (reference publishing, materials development, and language testing), the direct use of corpora in teaching (teaching about, teaching to exploit, and exploiting to teach), and further teaching-oriented corpus development (LSP [language for special purposes] corpora, L1 developmental corpora and L2 learner corpora)." (p. 364).

McCarten and McCarthy (2010) cited further work which also categorised the use of corpora in materials development into three approaches – (1) "corpus-driven" where the corpus provides the sole source for the language description, (2) "corpus-based" where the corpus instantiates examples of an existing language description and (3) "corpus-informed" which is a combination of the previous two approaches (p. 13).

Beyond the use of corpora for dictionary creation (for example, the COBUILD corpus was used to create the Collins Dictionary, Grammar book and English Course and Sinclair, 1986), Sinclair (1986) discussed the use of concordancing tools for textbook development. McEnery and Xiao (2011) reported on the use of the Cambridge International Corpus as a basis for content generation for the Touchstone book series. The rationale behind using a corpus was to ensure that the language contained within the English language learning text represented more authentic language in use rather than more traditional 'classroom' English. Language testing was reported to be facilitated through extraction of words, phrases or sentences as a basis for cloze tests and other types of language tests (McEnery and Xiao, 2011). McCarten and McCarthy (2010) reported on their further use of corpus data in the development of the Touchstone series. They stated that "extensive use was made of a corpus to inform not only the grammatical and lexical syllabuses, primarily through frequency lists and concordances, but also the topics and methodology, including models for presentation and practice" (p. 13).

Similarly, where corpora have been used to inform coursebook generation, frequency counts have been the predominant form of metric used. McCarten and McCarthy (2010) described how frequency lists must be used with caution, and frequency bands are a more

reliable method to use. Frequency bands, or blocks present content in groups according to frequency. Sometimes words from the same semantic group (e.g. days of the week or colours) can occur in different frequency bands, but will usually need to be presented to the language learner together (McCarten and McCarthy, 2010).

McCarten and McCarthy (2010) also reported on how they used corpus data to generate the conversation management syllabus in their coursebooks. While frequency lists were the starting point, it was noted that certain words like 'actually', 'absolutely', 'basically' and 'anyway' occurred more frequently in speech than in written texts. Frequency lists of multi-word items (also called "'clusters', 'chunks', 'lexical phrases' or 'lexical bundles'" (p. 14)) between two and seven words long were also extracted to be used as "productive stems" (p. 14) for language production. Examples of multi-word items were 'I mean', 'as a matter of fact' and 'I don't know if/what'. McCarten and McCarthy (2010) made the dichotomy between multi-word items providing "relational language and transactional language" (p. 16). The former is used to establish and manage relationships with others, while the latter is for carrying "real-world informational content" (p. 16). A combination of single word frequency lists, multi-word lists and comparing results from different corpora all contributed to isolating the relevant content for their language coursebooks. Compiling conversational language required the additional steps of examining frequent multi-word items in context and categorising these according to their language realisations – (1) organising your own talk (e.g. "I mean, on the other hand"), (2) taking account of the other speaker (e.g. "you know, things like that"), (3) listenership (e.g. "uh huh") and (4) managing the conversation as a whole (e.g. "speaking of, anyway") (McCarten and McCarthy, 2010, p. 17-19).

Teachers have also used corpora in direct language teaching and learning, where collocations and frequency lists have been put into play directly in the classroom. These facilities are, more often than not, suited to the more advanced learner where they can be asked to make implicit judgements on grammar, co-occurring words or single words.

Braun (2005) highlighted the difficulties around corpus use where their size, content, data format and annotation can have a negative impact on how appropriate they are for language learning means. She posited that these corpus features, which are geared towards linguistic analysis, can act as a barrier to language learners who may not be able to pinpoint the language required among the noise of the linguistic markup and scale of larger corpora.

Braun (2005) cited work on a number of solutions to scale and content which enable corpora to be more useful in language learning. These solutions include the use of sub-

corpora which comprise a subset of larger corpora, the "use of small genre-specific corpora … and corpora created by teachers and learners themselves" (Braun, 2005, p. 50). Braun (2005) and Braun et al (2007) cited the downside of such corpora where they have been described as 'ad hoc', 'home-made' and 'quick and dirty solutions' in the literature. She concluded that the "successful use of corpora for learning and teaching hinges to a great extent on a successful 'pedagogic mediation' between the corpus materials and the corpus users" (Braun, 2005, p. 61).

Braun (2007) reported on the use of the ELISA corpus (which contains video-based interviews with speakers of English) in a secondary school in Germany. She described how her study was only the second documented use of corpora in secondary school settings (rather than at tertiary-level education). Three "pedagogical requirements" were recorded with regard to this setting: (1) the need for "pedagogically-motivated corpus content", (2) the need for a curriculum-based corpus to allow for "contextualisation" and curriculum integration and (3) "pedagogical mediation of the corpus" (p. 308-9).

These three criteria echo the requirements for all teaching and learning resources. They provide some of the reasoning behind the need for corpora to be mediated by either lexical analysis/corpus retrieval software or a practitioner before they can be accessed by language learners as a learning resource. Corpora are not usually designed as a language learning resource although they have been used to great effect in language learning. They provide a source of authentic language for language learners, although the corpus register and level of language need to be borne in mind, they can be used to examine language in context and use, as well as for the discovery of language rules or patterns. The considerations emerging for corpora and for other language learning databases are similar to one another, as well as being similar to those emerging for paper-based resources like textbooks:

- Are they curriculum-based?
- Are they localised to the needs of a particular target group?
- Are they directly applicable to language learning activities or is there a need for the database/corpus content to be mediated (by software or the teacher/practitioner) for the language learner?

The next section examines content databases which have been developed. The systems described retain the content separate from its product so that it can be reused in multiple ways and settings.

3.2.5 Language learning databases: content for resources

Developing content for language learning resources takes much time, energy and expertise, regardless of the format through which it will be presented to the learner. Often the level of the requirement for these three assets is determined by the importance placed on content by the developer(s) or practitioner(s). Colpaert (2004) highlighted the importance placed on content development in CALL courseware development – "most of the work in terms of time and energy was spent, not on software development, but on content development, which for most packages took one" (p. 15). Decoo and Colpaert (1997) reported that 19/20 of the full-time staff at the former Didascalia were devoted to content "selection and organisation" (p. 74), with only one member of staff being responsible for programming.

McGrath (2002) noted that the time consuming nature of materials production militates against teachers developing their own resources. Similarly, Brusilovsky, Gamper, Knapp (2006) cited the time required to develop resources which also required "experts" (p. 199). Hutchinson and Waters (1987) warned not to underestimate the time required for writing materials. Content-based systems such as that reported on by Knapp (2004), can require many hours for the initial content to be authored, sourced and input.

Colpaert (2004) noted that often, language learning content can be tied up in the format through which it is presented, and difficult to extract in order to reuse it. He maintained that all content should remain separate from its display mechanism, so that it can be reused in alternative ways and formats, hence maximising the time/cost factor of content development. Ward (2002) and Gamper and Knapp (2002) have provided evidence of the successful separation of content or data from its application in language learning (be it CALL courseware or electronic learner dictionaries), so that content can be reused and applied in whatever format is applicable. Similarly, Decoo and Colpaert (1996, 1997; Decoo, 2011; Colpaert, 2004) have reported in different domains on the development of content databases for teaching and learning languages. These databases have been used as the content source for the development of textbooks and other language learning resources. These systems will be discussed in the coming sections, with a similar focus on the three challenges of content separation, content localisation and reusability of content. As separate content database systems, they already address two of the three challenges. The remaining one to be examined is their stance on the localisation of that content. They also start to build up a picture of *how* a solution to all three challenges can be reached through the development of content databases.

In some cases, the use of tools from the Computational Linguistics (CL) domain helped to automate some of the process(es). CL, also called Natural Language Processing (NLP), draws on linguistics and computing to process natural language using the computer. McEnery (1992) provides a good overview of the scope of CL, which includes areas such as document processing, spell checkers, thesauri, machine translation, text processing, natural language interfaces, style checkers and lexicography.

Many tools have been developed within CL/NLP for various languages. Most tools tend to be language specific and concentrate on majority languages. There are many tools for German (part-of-speech taggers, morphological analysers, speech systems etc.), but very few have been developed for Irish for example (although interest is growing – the Microsoft Irish spell-checker (Burke, 2002) and a Finite State Transducer (FST) morphological tool (Uí Dhonnachadha, 2002) are examples). Some of these tools have been used in Intelligent CALL (or ICALL) systems (e.g. Heift, 1998; Rimrott and Heift, 2005; Vandeventer, 2001) but the scope for application largely exceeds any use already made (Keogh, Koller, Ward, Uí Dhonnchadha and van Genabith, 2004).

The ELDIT system: semi-automatic generation of a content database using CL tools

Knapp (2004) reported on work on the ELDIT language learning program which involved an alternative way to author content. Knapp pointed to the time-consuming task of authors inputting language content for the ELDIT system, and how automatic computer programming processes were written to partly automate this process. As an initial stage of work, authors manually input language content in a semi-structured format. This content was then processed and marked up using resources such as databases, knowledge bases and tools from Computational Linguistics, such as generators and analysers. The resulting "detailed, structured and enriched data" (Knapp, 2004, p. 106) could either feed back for the author to edit via an authoring tool, or be used by the CALL program. XML (eXtensible Mark-up Language) was used as the content database mark-up language.

Knapp (2004) discussed the process of dictionary generation from manual input to final use. Content authors (who were linguists) used an XML template interface to input words manually. This was a labour intensive task. They also inserted some referenced characters as markers for emphasis, or to indicate prefixes or suffixes for example. Once the content has been entered, a series of CL tools, among other processes, are run on the content to expand and enhance it further. These tools included:

- Spell checker: to check the spelling and suggest corrections where necessary
- Lemmatiser and morphological analyser: to trim words down to their stem and add details about the morphological make-up of each word (e.g. prefix, suffix, plural)

- POS tagger: to add the relevant part-of-speech tag to each word to assist with word disambiguation and for future use.

(Summarised and extended (CL details) from Knapp, 2004).

A text corpus (as well as dictionary input) was also used to generate content, using the same process supplemented with a grammatical analyser or "classifier" (Knapp, 2004, p. 110). Content from both this and the dictionary development process was used to feed into the ELDIT CALL program. The program can be customised to German/Italian for learners of German/Italian, but it is largely a global product suitable for a large cohort.

Separation of content from processing

Ward (2002) also reported on the development of an XML-based database which fed into the generation of language teaching and learning materials. This project was developed to provide a written legacy for an endangered language, the Nawat language (Pipil), spoken in El Salvador, as well as to provide an opportunity for non-L1 speakers to learn their traditional language.

When it came to deciding on the content to be included or gathering content for the database, Ward described how the content was chosen by examining syllabi from other jurisdictions – "ideas about content are mainly drawn from a summary of syllabi from Washington State University (Callaway, 1985), Cambridge University (Swan and Walter, 1984) and ALLC (Clark, 1987), some adaptations had to be made" (Ward, 2002, p. 106). These sources all refer to foreign language learning courses and curriculums. The course content comprised 12 lessons all based on a series of templates with the learning objectives as follows:

- "become familiar with basic Nawat greetings and vocabulary,
- learn how to ask and answer basic questions,
- learn how to form basic phrases/sentences in Nawat,
- understand basic spoken Nawat phrases,
- read simple Nawat text."

(Ward, 2002, p. 180).

Once the relevant content was decided, Ward populated the database with relevant words, phrases and sentences in the target language (2002). Relevant image and audio files were added to complement the content and add cultural relevance. The content was stored in an XML database. This database fed into template-based display modules (XSL-based) which output HTML (webpage) pages to display the content including the multimedia elements (audio and images). Ward's system (2002) is a localised system, which considered

linguistic and cultural norms for its target learners and retained strict separation of data from its end product.

Enriched linguistic databases for textbook and courseware generation

Colpaert and Decoo have reported on work undertaken at the Linguapolis (formerly named Didascalia[9]) Research Centre since 1996. Linguapolis produced CALL materials for the "CALL industry, the Flemish school system and the private sector" (Decoo and Colpaert, 1996, p. 297). It is through a historical tracing of the works published by Decoo, Colpaert and their colleagues at Linguapolis, that a picture can be established of how their products are built up and developed over time. Their products represent a rich insight into how content databases can be developed and put to use in language learning resources.

Decoo (1996) outlined how theories of teaching and learning can rarely be incorporated to any great benefit within the classroom, as teachers do not have the time or means to "transform theoretical recommendations into efficient learning materials" (Decoo, 1996, p. 300). Decoo and Colpaert (1997) reinforced this point, and added how the reality of a busy classroom with many students and the demands of a national curriculum and/or examination to meet can prevent teachers in updating their teaching methodologies to take advantage of an improved methodology. As a result, he set about integrating these learning recommendations into a series of textbooks which had their basis in a content database. The key feature of this series was the "systematic build-up and practice of content over longer periods" (Decoo, 1996, p. 300). Further areas of work built up the idea of "external versatility" (van Elsen, van Deun and Decoo, 1991) which is a process whereby a large content base and accessible front-end of a CALL program make a CALL program usable in a large number of settings by a variety of user types. Table 3.6 below sketches the guiding principles and processes underlying work at Linguapolis.

[9] Linguapolis was set up by Colpaert and Decoo over two decades ago, under the original name of Didascalia. The research centre is very successful and carries out a wide range of activities from research projects, consultancy, organising conferences and presentations at conferences to writing publications and editing a leading journal in the CALL field (Colpaert, 2004). Many successful language learning programs for real-world users have also been developed here.

'Didascalia' will be used in instances where the reference predates the change in name to the company.

Foreign Language Learning	CALL
▪ Needs analysis of the target group, taking learning potential and individual learning environment into account ▪ Frequency and context studies for content generation ▪ Content structuring, presentation and organisation ▪ "Efficiency studies for didactic strategies": highest return yielded from any "learning 'energy'"	▪ Ensuring "external versatility" ▪ Organisation of content database – specifically around the concept of a didacteme (see definition below) ▪ Ensure ergonomic use of the computer ▪ Implement "degrees of freedom" in user interaction with the system ▪ Develop "testing modules" for CALL

Table 3.6 Principles and processes underlying work at Linguapolis (compiled from Decoo, 1996, p. 303-304)

Colpaert and Decoo (Decoo, 2011; Decoo & Colpaert, 1996, 1997) described their development of content databases as comprising of constructs they termed didactemes. They defined a didacteme as "the minimally definable learning entity" on which "a learner must concentrate for mastery" (Decoo and Colpaert, 1997, p. 68. Didactemes are described as items which have the same "distinctive features" (Colpaert, 2004, p. 193). Distinctive features "represent all of the information needed or parameters required for content selection, error analysis, tutoring and tracking in the program" (Colpaert, 2004, p. 193), that is, the metadata that is attached to content words, phrases and sentences to describe them (word category, tense, gender etc.). Didactemes are *combinations* of distinctive features, for example, "he eats" and "he sleeps" both have the same didacteme as they are both pronoun+verb combinations in the third person singular present tense. Each didacteme is therefore a unique descriptive set of metadata attached to database content.

Didactemes can be lexical, grammatical or cultural. In the case of a lexical didacteme, Decoo and Colpaert describe a construct which

> "can be defined for its formal and its structural appearance as well as for its semantic value. Each combination of a word with additions peculiar to it, such as specific prepositions, any idiomatic expression, or compound words, was identified as a didacteme" (1997, p. 68).

Where a word can have more than one meaning, for example, bank – as a financial institution, as the side of a river or as a verb – each form/use of the word would be presented as a separate didacteme, each with its own "additions peculiar to it" (Decoo and Colpaert, 1997, p. 68). Colpaert and Decoo have used content databases as the basis for textbook and CALL courseware content.

Decoo (1996) discussed the importance of didactic impact, that is, the extent to which the outputs of development work at Didascalia/Linguapolis matched the needs of the potential learners. He stated that "the didactic impact is very high when CALL matches textbooks or situations" (Decoo, 1996, p. 304). This same principle is at play in many domains where learners are exposed to the same content source through a variety of modes, so reinforcement and exposure to language are maximised.

Decoo provided a snapshot of the content database at play behind their early lexicon and verbal form systems – "content is not stored page after page, or lesson after lesson, as a chronological handbook would require, but in structures of minimal linguistic elements, each carrying a set of codes for retrieval according to the required learning situation" (Decoo, 1996, p. 305). He makes the case for building language from a bottom-up approach, starting at the word level. This approach avoids the pitfalls of prior top-down approaches to language processing and generation, which attempt to account for an entire language rather than build systems which can successfully account for a subset of language (built from the word level upwards), as seen through successful projects like the Météo weather forecast translation system in Machine Translation (MT), for example.

Decoo also merits the extensiveness of the content sources generated at Didascalia/Linguapolis as providing "richness of content" with over 1,300-1,500 verbs for their early verbal programs and 5,000-9,000 items for their lexical programs (Decoo, 1996, p. 307). Richness of content is important in providing variety for the learner, as well as a comprehensive source of content from which to work. Decoo, Heughebaert, Schonenberg and van Elsen (1996) described the process of how the content is selected and written for these systems:

1. ***Selecting headwords*** (words, idioms or structures) according to frequency (more often used words are generally given priority), utility (words and phrases to enable the learner to complete daily functions), coverage (words with metalinguistic qualities which can be used to describe other unknown words/concepts)
2. ***Integrating separate files into one unified file*** where an automatic process combines and ranks headwords according to (a) the importance placed on their source list and (b) the number of times a headword occurs across the lists.
3. ***Coding each headword*** to allow for retrieval for specific units. Coding includes (a) word class (part of speech), (b) semantic field such as 'diseases' or 'positive attitude' and (c) didactic unit where headwords are grouped together into 20-30 words relevant to frequency and semantic unit.
4. ***Adding information to each headword*** such as (a) semantic context or sample sentences which elicit the new word, (b) translation into the L1 of the learner to allow

for semantisation – the process whereby learners best learn new words with the" direct help of their translation in the mother tongue" (p. 332) followed by use in the target language, (c) linguistic comments such as polysemy or contrastive information and (d) synonyms

(summarised from p. 323-325).

Decoo et al (1996) reported on how semantic reply forms (SRFs) are used to add context for new words being learned. These SRFs are a combination of question and answer whereby the new word is the final word of the reply. "Synonyms, antonyms and collocations" are used as "semantic indicators" when they are being produced (p. 333). A sample SRF provided for 'rent' was:

- "Does Viktor have to leave his apartment?
- Yes, he can't pay the rent"

(Decoo et al, 1996, p. 332).

The end content source is presented to the learner in a menu-type interface where the learner can pick/choose how they would like to interact with the content. The content is presented *systematically*, so the learner builds up their repertoire gradually and over time, and encounters repetition and recycling of words so that known vocabulary can be extended with new ones (Decoo et al, 1996). New words presented in each sitting for the student are scaffolded by known words. However, words which are labelled for later encounter will not be presented until the relevant progress has first been made by the learner.

Heughebaert and Troubleyn (1996) outlined how the content generation stage differed in the development of a system for specific businesses' purposes, e.g. cockpit announcements or telecom terminology. The first stage of content generation involved gathering reports, internal documents, manuals and any other language sources relevant to the target industry. These sources were drawn together into an electronic corpus, which was then processed to arrange the content by frequency. This database arranged by frequency was then lemmatised, that is, each word was stripped back to its base form. These corpora can be extensive, and in consultation with the client, are usually stripped back to the 3,000 most relevant words. Some gaps in vocabulary are filled at this stage, as frequency does not always present the most important words in any one sector. Semantic fields are chosen in close collaboration with the client and the words are then evenly *spread* across these fields. Words are also *spread* over the different grammatical categories – nouns, verbs, adjectives and adverbs. Heughebaert and Troubleyn (1996) explained that in these types of programs, "the nouns make up 80% of the database" (p.352). Contextual sentences are then added (a processes which was reported to take nearly half of the development time), along with

synonyms and translations of the context sentences. The final additions to the database are codes which can be used to identify and call upon any unit of the database. These codes were reported to identify elements like word class, multiple choice word type and lesson code (Heughebaert and Troubleyn, 1996).

Decoo and Colpaert (1997) also charted the process of using a content database from which to generate a textbook. They described the following three steps, as summarised below:

1. *Selection of content based on didactic criteria*: Didactemes were developed to support the selection of didactically-based descriptions of words and structures to be taught and learned. This work was a combination of manual semantic reflection, and automatic processing. Each lexical didacteme was also manually assigned to a semantic field. The result of this phase was a refined database from which material would be selected.
2. *Organisation of content based on didactic criteria*: The refined database was then processed into frequency blocks (groups of words which occur with the same frequency), rather than just by frequency, to ensure that colourless and artificial texts would not be present when the most frequent words alone were presented. The frequency blocks allowed words to be presented from the same level but with different semantic groupings, morphological groupings and word lengths. Decoo and Colpaert outlined how "it is more efficient to present words within a semantic field by driblets, on a vast circle containing all the semantic fields. At each rotation, a few new words are added to each field" (1997, p.70). The importance of morphological groupings comes from the premise that "certain word categories are more difficult to learn, especially adjectives, adverbs, prepositions and conjunctions" (Decoo and Colpaert, 1997, p.70). Lessons should therefore not contain a higher proportion of these words than *in the total spectrum.* Variation in word length was reported to enhance learnability.
3. *Reiteration of what has been learned*: Recycling and revision of language (pre-knowledge) across a learning scheme/programme helps language learning to be consolidated. Computer processing was used to track which words were first introduced and when, so that they could be re-presented at the relevant time (and within the same semantic field).

(Decoo and Colpaert, 1997, p. 68-72).

More recently, Decoo (2011) has substantiated the importance of accounting for didactic sequencing in learning materials. He posits how this can be attained through the development of didactic databases which "systemize" the content to be learned as well as present content in the appropriate order of acquisition (Decoo, 2011). Decoo (2011) describes how "…systemization could help optimize input facilitation by providing a

"curricular order" that matches, as well as possible, both a natural order of acquisition and the most favourable didactic sequencing" (p. 151).

The resulting textbook, Éventail, was a huge success, being adopted in more than 80% of secondary schools within its first two years on the market. The underlying source database was localised to the particular L1/L2 of target learners and the target language being learned.

Table 3.7 below outlines the many processes which have been outlined as being carried out on Linguapolis content databases. As the references come from different sources of literature and as the exact details are not available in some instances, the second and third columns, *system* and *process*, suggest the automatic or manual processes which could be at play to generate the database content.

Content processing	**System**	**Process**
Perform Frequency count / blocks	Frequency count	Automatic
Select headwords	Based on frequency count	Automatic
Carry out lemmatisation or tokenisation	Stemming of words down to their root or splitting text at the word/character/token level	Automatic
Carry out morphological grouping / part-of-speech (POS) tagging	Morphological analysis tool / POS tagging system	Automatic
Add words to semantic fields	Manual documentation of relevant set of semantic fields and assignation of each word to its relevant field	Manual – by a linguist
Count word length (to ensure variety in length among words in the same frequency block)	Word count tool	Automatic
Add additional content		
Translation of the word/phrase	Automatic using dictionary tool or manual	Automatic or manual
Synonyms	Automatic using dictionary tool or manual	Automatic or manual
Semantic reply form (SRF) for meaning	Manually conducted without any reference points	Manual – by a linguist
Sample sentences	Automatically compiled from a corpus or manually compiled	Automatic or manual

Table 3.7 Processes carried out by Decoo and Colpaert (1996, 1997) on content and assumption of the system and processes involved

Content databases as the content source for language learning resources are not a new phenomenon. Indeed, databases, by their very nature will form a content source for any type of content-based system (and many CALL systems hold some or all of their content in some form of database). While the content may be processed or revisited at some stage

later in the process, none of these systems reported trialling the content as an entity in itself with learners before the remainder of the CALL system or textbook was developed to present the content in its finalised product. Content separation and reuse were inherent to these content database systems, and each displayed varying levels of content localisation.

The discussion now turns to the parallel thread that has already started to weave through the discussion through the use of CL tools for processing content. This parallel thread involves the processing and examination of raw linguistic content which has been abstracted from its product. The purpose of this examination is to establish whether raw linguistic content differs from its fully formed cousin embedded in its end product and what characteristics it exhibits.

3.3 Examining raw linguistic content: processing, analysing and characterising content

Some examples of how language learning content can be processed were outlined in the previous section. Processing content by moving around content or altering its layout can improve the access and adaption capabilities of other programs on the content (e.g. moving each token onto its own line to facilitate using the data within certain programs). Similarly, the addition of certain mark-up or annotations on raw linguistic content can make its content more accessible and adaptable by other programs. For example, the addition of part of speech tags would allow certain word classes to be isolated and accessed with ease. Much of these processes can be conducted by hand but there are also a host of programs available (from the CL and text/corpus analysis fields) which can be used to conduct some of this work automatically.

With the textbook a firm staple in schools, and taking on a greater role than just resource provider (e.g. curriculum and planning documents), an examination of their content becomes relevant in examining the content which they cover (complexity and sophistication). As with other types of data analysis, content analysis has traditionally been conducted in two main ways - quantitative analysis and qualitative analysis. Content analyses can reveal "pedagogical, psychological and epistemological positions" (Harwood, 2010, p. 8 citing De Posada, 1999, p. 425).

Within the domain of content for teaching and learning languages, Harwood (2010) described **quantitative data analysis** as "counting the number of references to a particular

topic / item, or identifying content categories and calculating the percentage of space devoted to each category" (p. 8). Harwood (2010) described **qualitative data analysis** as being "more overtly interpretive, seeking to uncover meanings and values transmitted by the materials" (p. 8). Harwood (2010) also cited Wasburn's (1997) description of qualitative analysis as "revealing cultural patterns and the focus of societal attention" (Harwood, 2010, p. 8 citing Wasburn, 1997, p. 473) as materials will tend to embody the author's own culture and society.

The following sections examine the processing, analysing and characterising processes which can be carried out manually or automatically on language learning resources and raw linguistic content. These processes stem from the language materials domain, SLA and CL or MT, and are examined in each of their contexts through the coming sections.

3.3.1 Analysis undertaken in the materials writing and development domain

Harwood (2010) emphasised how content analysis can examine whether the language used in resources is likened to language which is used by native speakers. He cited Sheldon (1988) who found that publishing houses often selected vocabulary and content "without system" (Harwood, 2010, p. 9), such as consulting service lists which present the most frequently used words in a language and provide the learner with between 80% and 95% coverage through knowing them. Service lists can be up to 2,000-3,000 words long (e.g. West's 1953 list, or the more recent lists from Browne, Culligan and Phillips (2003) and Brezina and Gablasova (2003)) and require updating over time to ensure they take account of changes in speech and emerging words. Service lists are compiled from corpora.

Content analysis can reveal how the language used in materials is at odds with what is actually used in naturally occurring language. Arnold and Rixon (2008) described how the language presented in young learners' textbooks did not match language which is usually used by young native speakers. Similarly, Harwood (2010) presented studies which revealed how linguistic areas such as modal language, reported speech and pronunciation have been misrepresented in materials. He also accounted for variation between authentic and textbook materials in different genres like law and business (Harwood, 2010).

Smiley and Masui (2008) analysed the English language found in textbooks produced by three different sources – (1) globally-produced materials, (2) materials produced by native English speakers for the Asian market and (3) materials produced for Japanese publishing houses. They used the Mann-Whitney test of medians to examine the total number of

activities, the number of pair-work activities (to gauge the level of communicative activities) and sentence length in texts from texts in categories (2) and (3). They outlined that the "Mann-Whitney test is useful for making inferences about the difference between two population medians where the sample sizes are small and do not have enough data to produce a standard bell-curve" (Smiley and Masui, 2008, p. 247). The number of relevant counts in each case was compiled manually and then transferred to computer for calculation (N=14 texts from each category, and N=38 phrases from one textbook source). They concluded that there were fewer activities in type (3) texts as Japanese teachers generally use a textbook from start to finish rather than picking and choosing activities which would be the norm for type (1) and (2) texts. They also concluded that some type (2) texts were a little too basic in their demand on students with short sentences and basic vocabulary (Smiley and Masui, 2008).

Harwood (2010) also outlined how analysis can reveal how materials can be culturally inappropriate (e.g. areas such as religion, politics and alcohol are generally avoided) and present unrealistic stereotypes. Littlejohn (2011) proposed a framework for the analysis of content of published materials. Littlejohn (2011) noted how published materials were so widespread and how they have "evolved into much more complex objects…offer[ing] complete packages for language learning and teaching" and "effectively structure classroom time" (p. 190). With this in mind, Littlejohn developed a framework for analysing materials, to ensure that they were suitable for the latter demand.

"1. Publication
- Place of the learner's materials in any wider set of materials
- Published form of the learner's materials
- Subdivision of the learner's materials into sections
- Subdivision of sections into subsections
- Continuity
- Route
- Access

2. Design
- Aims
- Principles of selection
- Principles of sequencing
- Subject matter and subject of subject matter
- Types of learning/teaching activities: What they require the learner to do; manner in which they draw on the learner's process competence (knowledge, affects, abilities, skills);
- Participation: Who does what with whom
- Learner roles
- Teacher roles
- Role of materials as a whole."

Table 3.8 Littlejohn's framework for content analysis of published materials (Littlejohn, 2011, p. 183)

Littlejohn (2011) provided for three levels of analysis, each of which indicated the level of difficulty in accessing the information required. Level 1 for example, was easily ascertained information on the published form of the learner's material. Level 3 was more involved, requiring manual analysis of much of the book before the relevant information could be compiled (e.g. aims and principles of the publication where they are not stated by the coursebook author). All of the areas of analysis were to be carried out by hand through analysing the materials.

Most of the studies mentioned above relied on the manual counting of relevant items, which were then sometimes transferred to computer for compilation and presentation. This could be the case as textbook content is hard to manipulate automatically when it is embedded amongst its published layout mark-up and images, as well as being locked into more inaccessible formats such as PDF. Other studies have started to integrate more automatic means of analysing textbook content through using corpus analysis tools on corpora comprising language from language learning textbooks.

Mukundan (2010) reported on such work which drew on automatic analysis of content. He examined two Malaysian textbook corpora – "The Corpus of Textbooks of the 1998-2003 Cycle of Malaysian Secondary School English Language Textbooks … The Corpus of Textbooks of the 2004-present Cycle of Malaysian secondary School English Language Textbooks" (Mukundan, 2010, p. 292) – for repetition and recycling of vocabulary. The number of tokens (all words in a text) and types (the different words in a text, so each new token is counted only once) were used (through the use of WordSmith software) to ascertain the frequency of occurrence of vocabulary items to examine whether new vocabulary was presented often enough for retention (on the premise that it was required seven times to allow the new vocabulary to be assimilated and remembered) (Mukundan, 2010).

3.3.2 Analysis undertaken in SLA: complexity, accuracy and fluency (CAF) and error analysis

This section introduces the concepts of text analysis through an SLA lens. The focused viewpoint of SLA outlined in this section examines only those areas which are most relevant to the text analysis work undertaken in this research. Broader considerations such as SLA approach, teaching method and assessment method are separated out in this instance, as described through the introduction to raw linguistic content in Chapter 1 and further defined in Section 3.1. Section 4.1.4 describes additional areas of SLA focus through the background to the localised settings which are relevant to the local context in this research. These are presented according to the localised settings of the Primary School

Curriculum (PSC), curriculum linkage and integration, the modern language curriculum documents, language awareness and learner error analysis. Section 4.1.4 describes the teaching approaches and methods espoused in the Irish Primary School Curriculum and complementary draft modern language curriculum, as well as delving into the underlying processes which can sometimes account for learner errors.

The SLA heading has been retained in this present section, as it represents a common overarching domain for the introduction to Complexity, Accuracy and Fluency analysis (CAF) and error analysis studies. While error analysis is exclusive to the SLA domain, analysis using CAF ratios can cross discipline boundaries, as outlined in Chapter 1. Malvern et al (2004) described some of the related areas of "linguistic investigation" (p. 6) which integrate measures of lexical diversity (complexity) for instance, such as in the analysis of historical documents, studies of authors' style and dating works of literature, forensic linguistics and diagnostic linguistics.

Complexity, fluency and accuracy

Qualitative and quantitative content analysis can reveal different features and characteristics about local learning preferences or the suitability of content to its targeted learners and curriculum. SLA offers more means through which quantitative analysis can be carried out. Traditionally employed by teachers who wanted to track the progress made by their learners, studies from SLA present different ratios and matrices for analysing content for complexity, accuracy and fluency (CAF).

Wolfe-Quintero, Inagaki and Kim (1998) outlined how measures associated with "fluency refers to speaking with 'native-like rapidity', accuracy refers to being 'error-free' and complexity refers to 'using a wide range of structures and vocabulary'" (Wolfe-Quintero et al, 1998, p. 4).

These measures are used in SLA studies to measure *language development* in learners. Language development refers to the characteristics of a learner's output which can reveal some point along a developmental continuum. A developmental index can also be developed and used as a scale which would allow learners to be compared to one another and across studies. *Writing proficiency* can be examined as a subset of language proficiency, but is a broader term to language development.

An examination of the complexity of language looks at the range of structures and vocabulary used within it. Hence, syntactical or grammatical complexity refers to structures and lexical complexity refers to vocabulary. Varying measures have been used to examine

complexity (Vercellotti, 2012; Wolfe-Quintero et al, 1998), starting with the base level of the proportion of types to tokens (see Table 3.9) which would indicate the variation in vocabulary used (and hence, the complexity of the language used). The alternative formulae have been outlined to try to best account for variances in text length. The type-token ratio is the most basic of these, and simply calculates the proportion of types (unique words) to tokens (words). This ratio can present varying results depending on the overall length of the text. The standardised TTR (STTR) ratio calculates the same ratio over certain lengths of texts, and then calculates the mean TTR over all of these individual ones to form a result which is less variable on texts of different length. The remaining three ratios all take this latter idea and present alternative ways of discounting the impact of text length on the ratio.

Ratio	Formula	Description
Type-Token Ratio (TTR)	$\frac{Types}{Tokens} \times 100$	Types: Unique words Tokens: Words
Standardised TTR	Measures the TTR but takes account of the total text length by taking the mean TTR across certain text intervals (e.g. if n=500, the TTR is calculated every 500 words and then the average TTR taken as the STTR.)	
Mean Segmental TTR (MSTTR)	Measures the TTR but takes account of the total text length by taking the mean TTR across text segments of 10 or 50 words in length	
Guiraud adjustment of TTR (2012)	$\frac{Types}{\sqrt{Tokens}} \times 100$	Types: Unique words Tokens: Words Adjustment to account for text length
D	Computes the TTR over a random selection of 35-word units per 100 words. D is the mean TTR of 100 samples of random 35-word units (words are randomly selected only once across the 100 samples). This should be repeated for 36-word segments, 37-word segments etc. up to 50-word segments (Koizumi, 2012).	

Table 3.9 Varying formulae to compute the lexical complexity/density of a text

Further complexity measures examine the level of complexity of varying word classes – verb variation, noun variation and adjective variation. Differing ratios have been proposed for these same counts (e.g. see Table 3.10 below), which all require identifying the relevant parts of speech in a corpus/text and counting them to compare their proportion to the total number of words or subcategories of words (e.g. lexical words which would comprise nouns, adjectives and verbs but exclude auxiliary and modal verbs) (Wolfe-Quintero et al, 1998).

Ratio	Formula	Description
Verb Variation 1	$\frac{VT}{V}$	VT: Verb types, V: Verb tokens
Verb Variation 2	$\frac{VT}{LW}$	VT: Verb types, LW: Lexical words tokens (Lexical words: nouns, adjectives and verbs)

Table 3.10 Varying formulae to compute the verb variation of a text

The next level of complexity analysis examines the proportion of sophisticated words present in a text. Sophisticated words are those which are not generally in listings of the most commonly used words (compiled from frequency listings). These ratios again progress to examine the sophistication of different word classes. The reverse of this process is to examine the 'basicness' of a text, which counts the proportion of words in a text which are in the most frequently used words listing.

Aside from their traditional uses by teachers to examine learner language for competence, and progress, they can also be employed to compare different types of texts to one another – specifically, raw linguistic content, textbook content and authentic language learning content (from newspapers for example). This type of analysis is the focus of Chapter 6, where the prototype raw linguistic content is compared and contrasted to other types of language leaning content to ascertain their makeup.

Error Analysis

Learner error analysis is a further type of analysis which stems from the SLA domain. Its relevance to the research lies in its use to add a further level of localisation to language learning materials. Analysing learner errors has been the method employed by teachers and researchers to predict a learner's level of ability in the TL (Lightbown & Spada, 2006), to analyse whether any progress has been made in language learning over time (Wolfe-Quintero et al, 1998), as well to predict typical errors which may be made by learners (Heift, 1998; Heift and Schulze, 2007). Error analysis has also been used to understand how learners "process second language data" (Lightbown & Spada, 2006, p. 80), as well as to develop computational models to predict, identify and even correct learner errors (Heift, 1998; Granger, Kraif, Ponton, Antoniadis and Zampa, 2007). These computational models are referred to as error tagging systems, and are usually associated with learner corpora. Learner corpora are collections of electronic texts comprising language produced by learners (written text or transcribed speech). Granger (2003) described learner corpora as "electronic collections of authentic foreign or second language data" (p. 465). Granger et al (2007) outlined how the term "error corpus" (p. 254) is sometimes also used to describe a learner corpus, although learner/error corpora will contain correct formations as well as

incorrect ones. Granger (2003) cited the benefits of learner corpora as (1) being computerised which means that the data stored within them is available for electronic linguistic analysis and manipulation and (2) their size which tends to be much larger than data sets traditionally gathered and used by SLA researchers (Granger, 2003). Granger et al (2007) outlined how learner errors can be analysed using a learner corpus by (1) using "a corpus-based method" by searching for words, phrases or sentences which the practitioner knows are causing problems or (2) using "a corpus-driven method" which compares learner corpus data to native corpus data (p. 255).

Learner corpora are generally marked up with error tags which identify the errors made by the learners throughout the learner corpus. Granger et al (2007) described how "error tagging systems differ in the types of error taxonomy used, the granularity of the tagsets, the formats in which they are encoded and the possibility of introducing one or more levels of analysis" (p. 256). Granger et al (2007) described the three-tiered error tagging system employed in the FreeText project which describes the (1) "error-domain" (e.g. lexis, grammar), (2) "error category" (e.g. gender, tense, number) and (3) "word category" (e.g. noun, verb, adverb), as well as suggested corrections (p. 256). Other projects have reported the use of alternative error taxonomies, from single tier (e.g. Mackey, Gass & McDonagh, 2000) to double (e.g. L'Haire, 2007) and on to triple tier, as Granger (2007) described. Other taxonomies have focused on inserting a correction for an error, rather than describing it (Díaz-Negrillo & Fernández-Dominguez, 2006). Some taxonomies comprise a combination of two and three tier taxonomies, such as Thouësny's hybrid taxonomy (2011) which drew on the format of three other taxonomies (Granger, 2003; L'Haire, 2007; Mackey, Gass & McDonagh, 2000).

Learner error analysis within localised contexts can reveal typical errors which learners make and can be accounted for in language learning resources. Learner error analysis is relevant to the development of the prototype linguistic content database, which takes account of learner errors in its localised context. Learner error analysis is explored further in the local research context of the primary modern language classroom in Ireland in Chapters 4 and 5.

The next section examines further means of analyses which stem from CL and MT.

3.3.3 Analysis undertaken in the Computational Linguistics and (Machine) Translation domains: controlled languages and limited domains

Controlled languages and limited domains are both subsets of language. They have been used extensively in CL and predominantly in (Machine) Translation to improve the output of processes from these domains. Controlled languages are also used in the software localisation and technical writing industries to assist in the writing and simplification of text.

Controlled languages and limited domains are bounded subsets of languages. It is through the restrictions or ring-fencing of allowable language and structures that simplified and more predictable language emerges. More predictable language input to Computational Linguistics and Machine Translation systems results in better outputs from them. The coming sections examine controlled languages and limited domains. They describe how they are constructed, how to test for them and what they comprise of. Raw linguistic content is similarly ring-fenced by its associated localised features, such as a curriculum. Chapter 6 describes how this content is analysed to ascertain whether it exhibits typical features of controlled languages. This section concludes with the implications of controlled languages and limited domains on automatic processes such as machine translation.

Controlled Languages

Controlled languages are "a form of language usage restricted by grammar and vocabulary rules" (Arnold, Balkan, Meijer, Lee Humphreys and Sadler, 1994). They are '*forced*' subsets of language. A controlled language fragment is achieved by simplifying and shortening sentences, omitting redundant words and avoiding elliptical constructions, for example. O' Brien (2010) summarised controlled language rules as specifying "constraints on lexicon, grammar and style with the object of improving text translatability, comprehensibility, readability and usability" (p. 143). Controlled languages have been used to great effect in Machine Translation (MT) (Arnold et al, 1994) and are usually implemented by passing input through a critiquing system or controlled language checker before it is processed. Controlled languages are used by large companies seeking to simplify their technical manuals such that they are accessible to non-native speakers of the document language and non-experts in the specific field (Kittredge, 2003). One of the best-known examples of a controlled language is AECMA Simplified English, an international standard for writing technical manuals in the aerospace industry (Kittredge, 2003).

There are many examples of controlled languages, such as AECMA Simplified English, Attempto Controlled English, IBM's Easy English, Sun Microsystem's Controlled English, ASD Simplified Technical English and TETRIS 2002 project CL (German) (Kittredge, 2003). Some controlled languages are developed to assist human-oriented processing (HO). The controlled language rules in this instance improve the readability, clarity and comprehensibility of texts. Those developed for machine-oriented (MO) processing improve the translatability of texts by making them more suitable for translation memory software and machine translation software. The Controlled Language definition depends on the intended use of the language adhering to the rules

An example of language rules set out for controlled languages is outlined in Table 3.11 below. These would be used in combination with a CL dictionary/dictionaries of allowable words.

Perkins Approved Clear English (PACE) Writing Rules	
Keep it short and simple	*Make it explicit*
▪ Keep sentences short ▪ Omit redundant words ▪ Order the parts of the sentence logically ▪ Don't change constructions in mid-sentence ▪ Take care with the logic of *and* and *or*	▪ Avoid elliptical constructions ▪ Don't omit conjunctions or relative pronouns ▪ Adhere to the CL dictionary ▪ Avoid strings of nouns ▪ Don't use –ing unless the word appears thus in the CL dictionaries

Table 3.11 The PACE writing rules of the Perkins Engines Engineering Company for achieving controlled languages (Kittredge, 2003)

Controlled language checkers are generally used after an author has written their text adhering to their controlled language. As the controlled language dictionaries can be large, the checkers are important in speeding up the writing process so the author can automatically check their content rather than referring to the manual time and time again (Kaiser, 2013). All of the well-known and large controlled languages have associated controlled language checkers (e.g. Boeing have made a controlled language checker for Simplified Technical English available).

In examining the PACE Writing Rules above, the types of analysis which could be associated with a controlled language checker would include an examination of the areas listed in Table 3.12 below.

▪ The number of tokens ▪ The number of types ▪ The type/token ratio (TTR) ▪ The Standardised TTR (STTR) (TTR which accounts for text length) ▪ The Average word length ▪ Average sentence length ▪ Number of punctuation marks ▪ Number of co-ordinating & subordinating conjunctions	▪ The number of relative pronouns ▪ A frequency comparison with 100 most freq. Words in German ▪ An examination of inflection: verbs, adjectives, articles ▪ An examination of the number of tenses used ▪ An examination of the use of the active vs. passive voice ▪ An examination of the indicative vs. subjunctive mood

Table 3.12 Sample features of controlled languages which can be examined through a controlled language checker

The features above would indicate the level of complexity in each text being analysed. For example, if a sentence contained more than one punctuation mark (the final punctuation mark to indicate the end of the sentence), then it would be more complex than a sentence which only contained one punctuation mark as it would likely include a conjunction and additional phrases. In the same way, higher numbers of conjunctions and relative pronouns in a sentence would indicate greater complexity than lower numbers of them.

The aim of controlled languages is to reduce the complexity of a text. As such, they have been used outside of multinational companies and also in educational settings to account for the readability of a text. Two formulae are used to calculate the readability of a text – the Flesch Reading Ease and the Flesch-Kincaid Readability Test (O'Brien, 2010) (Table 3.13). The Flesch Reading Ease scale is a 100-point scale; the higher the number, the more readable a text is. The Flesch-Kincaid Grade Level Test is based on grade level reading standards in the USA (Microsoft, 2014). A result of 0.6 would indicate that a text was suitable for a 6th grade student to read. MS Word and Outlook integrate these readability tests into their readability checker (Microsoft, 2014). This paragraph has a Flesch Reading Ease score of 46.8 (standard texts have a score of between 60 and 70) and a Flesch-Kincaid Readability Score of 10.6 (standard texts have a score of between 7.0 and 8.0) (Microsoft, 2014).

Ratio	**Formula**	**Description**
Flesch Reading Ease	206.835 – (1.015 x ASL) – (84.6 x ASW)	ASL: Average sentence length ASW: Average number of syllables per word
Flesch-Kincaid Grade Level Test	(0.39 x ASL) + (11.8 x ASW) – 15.59	

Table 3.13 Sample readability ratios and formulae

Limited Domains

Limited domains or sublanguages, are a subset of a language which occur '*naturally*' (as opposed to the *forced* subsets controlled languages represent). MT systems are designed for particular sublanguages like the language for hotel bookings or weather reports; they are limited to these domains and only cover the language which occurs within them. The Météo system is an example of the success of a sublanguage-based approach within CL/NLP. Météo translates weather forecasts bi-directionally between English and French (Hutchins and Somers, 1992). The 'weather' sublanguage has a small vocabulary, uses a telegraphic style of writing and omits tense. A sub-language restricts the vocabulary and constructions which can occur in input. It is then possible to anticipate and hence account for most, if not all input (Kittredge, 2003).

It is more difficult to test for limited domains, as generally, the test will compare any one dataset with a language dictionary containing vocabulary from a specific domain. The greater the co-occurrence of words between the specialised dictionary and the dataset, the greater the probability of the dataset being limited to the domain outlined in the dictionary. Any vocabulary occurring in the dataset which did not appear in the dictionary would also need to be counted and accounted for to ensure that the domain is not mixed. A longer text could contain a relevant number of tokens from the dictionary without necessarily conforming to being a limited domain corresponding to the dictionary content.

Implications of controlled languages and limited domains on automatic processing of text

CL/NLP systems can fail in coverage, robustness and/or accuracy. Given a particular application, CL/NLP systems must provide adequate language coverage, be able to process (mildly) ill-formed language input and deliver accurate processing results and responses. Recent research has exhibited how coverage can be addressed in Machine Translation, for example, by bypassing rule-based machine translation systems (RBMT) (which are produced by hand) and use Example-Based Machine Translation systems (EBMT) and Statistical Machine Translation systems (SMT) (van Genabith, 2009). These latter types of system can perform on larger data sets as they can account for a greater set of language.

There are other instances where the challenges of coverage can be addressed and provide full coverage, remain robust and be accurate (a dichotomy of CL/NLP-type systems which considers these three issues is presented in Keogh, Koller, Ward, Uí Dhonnchadha & van Genabith, 2004). In these instances, subsets of language, such as controlled languages and limited domains, can be used to obtain better results from CL/NLP systems. These subsets of language are predictable and limited so that the CL/NLP system can fully account for

their content. They restrict the input to a CL system suitably, so that the CL system need only handle the language and constructions which occur within the subset(s) of language.

One further example of CL/NLP systems overcoming the challenges relating to coverage are evidenced in their use in Intelligent CALL (ICALL) systems which have been successful in being implemented and used in educational settings. These are few in their number and include Heift's *E-Tutor*, Nagata's *Robo-Sensei*, Hagan's *Spanish for Business Professionals* and Amaral and Meurers' *TAGARELA* (Amaral and Meurers, 2011). They have successfully overcome the challenges relating to coverage by constraining learner input to the system through ensuring strict activity-types or formats where the input can be well-predicted (Amaral and Meurers, 2011). Amaral and Meurers (2011) describe how CL/NLP systems in this instance can be "tractable and reliable" when learner input "ill-formed or well-formed" are restricted (p. 11).

Providing evidence of the nature and characteristics of raw linguistic content for language learning, and whether it exhibits features of controlled languages, would have important implications for the use of CL/NLP systems on this content. Analysis matrices from SLA feed into characterising the raw linguistic content and providing evidence of controlled language and limited domain features (e.g. complexity measures).

The next section draws together the challenges and the problem space which has emerged across Chapters 2 and 3. It considers the global context for resources, their localisation, their reusability and the nature of language learning content.

3.4 Defining the global problem space

Having reviewed the status quo of resourcing in Chapter 2, and the development and analysis of language learning resources and content in this chapter, the discussion now focuses on the implications of the challenges identified and the requirement for a solution.

The first area for consideration is addressing the main resource challenges which were identified in Chapter 2 for consumable resources. These included how resources are time-consuming to create, the dominance of textbooks in place of curriculum-based planning and teaching and the requirement for localised resources. **Reusability of content** maximises the time/effort employed in their production as created context can be reused in alternative ways and settings. Similarly, retaining the **separation of content from its product** also allows for reusability of content. The **localisation of materials** ensures that they are relevant to their target learners and provide for their individual curricular and

learning needs. Discussions surrounding the importance of sustainability within the CALL domain, reinforce the validity of each of these issues in their interpretation of what 'sustainability' means. Blin et al (2013) note the importance of sustainability in CALL design and development which ensures the goals and results of CALL projects are maintainable beyond the lifecycle of the product under development. Jalkanen (2013) similarly refers to sustainability in CALL as the requirement for CALL systems to be adaptable and future-proofed so that efforts expended in their creation retain their value into the future. Liou et al (2013) refer to sustainability as the ability for work in CALL to be transferrable to other similar contexts.

The development processes for consumable resources indicate that many levels of abstraction can be generated between the author developing content for a textbook, and its final interpretation by a teacher in the classroom. The many development processes examined illustrate how varied and inconsistent the development processes are for language learning resources and content, and how these have had varying levels of consideration for the three main challenges identified for resources. Relevant development processes can be drawn from all of the processes examined to explore the development of a curriculum-based and localised linguistic content database.

	Challenge	**Potential solution and benefit**
Challenges surrounding resources and textbooks	1. Resources are time consuming to create	***Make resources accessible and usable*** - reduce the time it takes to create/draw together tailored resources
	2. Textbooks are used to replace planning/as a source of standardisation	Plan according to curriculum documents then source relevant resources. Ensure resources are curriculum-based
	3. Requirement for localised and customised materials	***Make resources suitable for local/individual teaching and learning settings*** – allow the teacher easy access to a source of relevant and tailored (cultural, level) materials
	4. Lack of resources	Make relevant and tailored resources easily available to the teacher
Challenges relating to resource development	5. Levels of abstraction between the curriculum and the learner	***Bypass the intermediating processes and people*** by producing curriculum-based resources which the teacher can access directly
	6. Lack of a robust and more scientific method of content authoring with learner trialling	Ensure content is developed in a robust way through trialling with learners in the target setting
	7. Localisation required for local resources	***Ensure resources are gathered and compiled to meet local teaching and learning needs***

Table 3.14 Summary of challenges identified

The solution envisaged to these challenges requires an examination of raw linguistic content creation for language teaching and learning resources. This can be housed in a

linguistic content database to ensure the separation of content from its end product(s). If content can be abstracted from its product, it can be developed as an entity in itself and then reemployed in various products (reusable). The development process of such raw linguistic content can aim to address the challenges relating to curriculum-closeness and localisation.

Figure 3.11 below summarises the main issues which are to be examined through this research. The research 'inputs' are those challenges which have been outlined above. The contextual requirements and parameters (e.g. localised settings) will be explored in the next section and will be specific to any one target research space. They are connected with the localisation of content to specific contexts. The research 'processes' are those which are required to reach the research 'outputs.' In this case, a new development model was required to guide the development of raw linguistic content and associated localised settings to produce a content database. The features of the research 'outputs' are also outlined. The output products are not included in the figure as they are determined by the teacher/practitioner. The output products are resources which draw on the content database for their content.

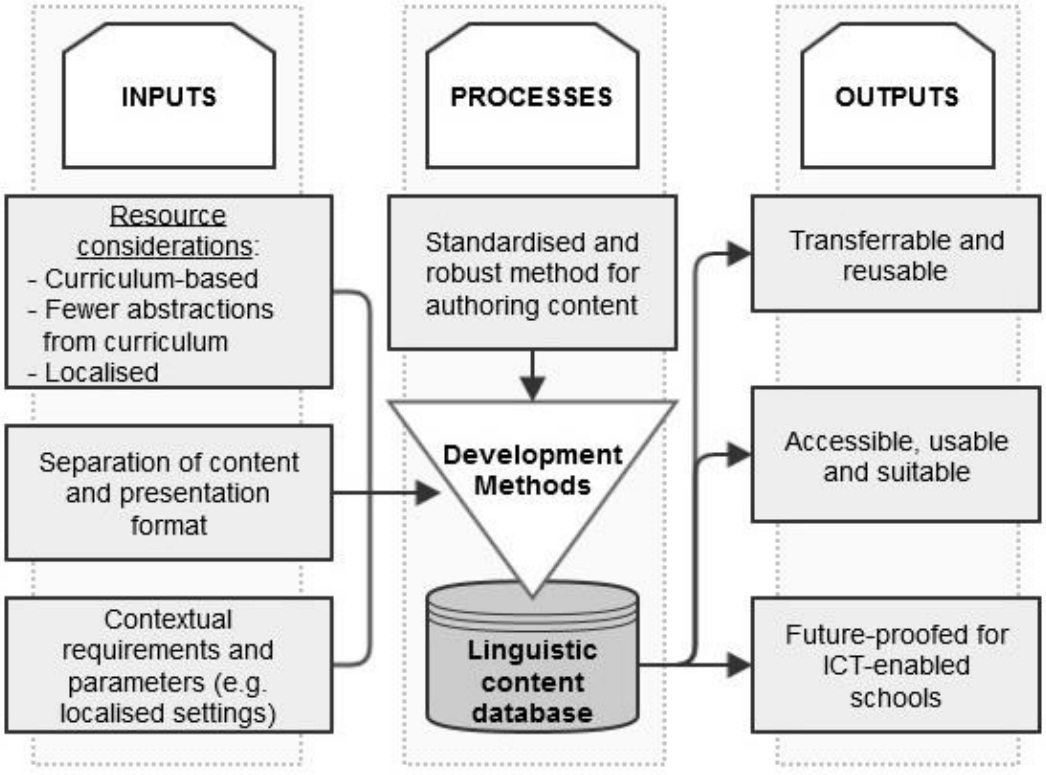


Figure 3.11 Summary of the problem space: key features of the research inputs, processes and outputs

One final area for investigation remains to be included in the research space. This is the processing, analysis and characterisation of raw linguistic resources, and their comparison

with other types of language learning content (Figure 3.12). These processes are carried out to ascertain the similarities and differences between the raw linguistic and the content from other language learning resource types and to characterise the raw linguistic content to ascertain whether it has features of a controlled language and limited domain.

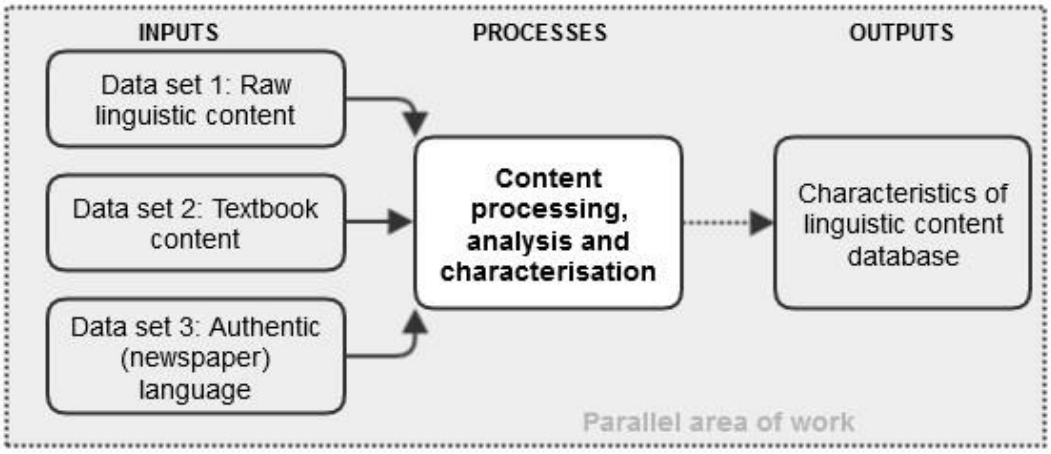


Figure 3.12 Parallel area of work: processing, analysing and characterising raw linguistic content

3.5 Summary and conclusion

This chapter commenced with a definition of raw linguistic resources. It focused on how teaching content is authored and created. Many educational disciplines consider these processes, and create them in a number of ways. Emerging development/writing methods can be divided up into five distinct groups, as those which

1. rely on intuitions, experience (as teachers) or their own creativity, write in an ad hoc and spontaneous way (Prowse, 2011; Fox, 1998; Tomlinson, 2008, 2012, 2013; Maley, 2013; Harbison, 2008)
2. use templates for lesson materials generation (rather than for an entire textbook or course) which specifically produce activities based around a text (Scott, Carioni, Zanatta, Bayer and Quintanilha (1984) and Axbey (1989) as cited in McGrath, 2002)
3. employ a framework for materials development (Tomlinson, 2013; Jolly and Bolitho, 2011)
4. follow guiding principles to author language teaching and learning materials (Bell and Gower, 2011; Tomlinson, 2011; Farr, Chambers and O'Riordan, 2010) or to ensure the integration of certain desirable features (e.g. authenticity, learner autonomy and cultural awareness as in Fenner and Newby, 2000)
5. use data-driven approaches such as employing corpora, concordancing and frequency lists (Sinclair, 1986; Farr, Chambers and O'Riordan, 2010; McCarten and McCarthy, 2010).

The focus of discussion also included an examination of whether any of the elements of the models accounted for the three resource challenges of reusability, localisation and separation of content from product, so that relevant processes could be pieced together to start to form a solution to the problem space identified. In all cases, the process of creating content is a resource heavy one which requires much time and labour. Content should therefore be made available for use beyond its immediate environment or purpose.

Section 3.3 examined how content can be analysed from the materials writing domain, the SLA domain and the Computational Linguistics and Machine Translation domains. These analysis matrices are put to use in Chapter 6 where the raw linguistic content is analysed and characterised.

This chapter concluded with a definition of the problem space and the parallel work which examines the raw linguistic content. It highlights how a more scientific and robust method is required for creating resource content, which ensures that learning content stays true to its educational focus and is localised to local teaching and learning needs. Content created in this way should be accessible to practitioners in a flexible way, requiring little time to access for employment in the classroom. At the same time, content should be made available in an adaptable way, to allow for its current and future use in whatever format is relevant to the educational setting (be it paper-based resources, or virtual/electronic ones). Finally, content should be made reusable, so that it can be adapted and employed in other relevant local settings.

The next chapter outlines the local context for the research, the research design and the methodology employed through each of the four phases or work.

Chapter 4 Research Context, Design and Methodology

This chapter examines the methodological implications of Chapters 2 and 3 in developing a reusable linguistic content database for resource generation. It examines how a reusable database of raw linguistic content could be specified and developed to address resourcing needs in a niche context. It explores the context for the research – the primary modern language classroom in Ireland, and its resourcing needs which have a global resonance as well as individual localised requirements. This chapter traces the research design and the methodologies used to progress through four phases of development – Phase I: Preparation for and research in the Irish primary modern language classroom and database specification, Phase II: Database development and population, Phase III: Database analysis and Phase IV: Database proof of concept application.

4.1 Research context: The Irish Primary School classroom

The Irish primary school environment has many resourcing challenges, many of which were outlined in Section 2.2. The majority of these challenges fit within the global themes identified and summarised in the previous section. English and Gaeilge/Irish are established language subjects of the Irish PSC. Four modern languages–French, German, Italian and Spanish–were part of a national Modern Languages in Primary Schools Initiative (MLPSI) between 1998 and 2012, which integrated modern language teaching and learning into participating project schools. The MLPSI fell victim to government recessionary cutbacks and the Initiative ceased in 2012. During its final year, the MLPSI team supported 546 schools which catered for 27,508 children (Kildare Education Centre, 2012). This represented approximately 17% of primary schools in Ireland, or 23% of children in fifth and sixth class in primary school.[10] However, within its 14 years in operation, the MLPSI allowed a wealth of research and data to be gathered on beginners' modern language teaching and learning in Irish primary schools which remains relevant to today's work on the development of a more integrated primary language curriculum (NCCA, 2014) and any future endeavours to integrate modern languages at primary level. Teaching and learning on the MLPSI was also grounded in curriculum documents – *Pilot Project*

[10] The Department of Education and Skills supported 3,159 primary schools (excluding special schools) and 509,040 children (excluding special schools) during the 2011/2012 school year – 119,610 of whom were in fifth and sixth class (Data sourced from DES statistics database, 2012).

on Modern Languages in the Primary School: Draft Curriculum Guidelines (NCCA, 1999) and *Modern Languages in Primary Schools. Teacher Guidelines* (NCCA, 2001).[11]

This modern language primary school environment provided the research context for the development of a prototype database comprising curriculum-based raw linguistic content and associated localised settings or parameters. Localised settings allow linguistic content databases to be localised and tailored to any one context, and to provide the flexibility for relevant content to be reused in other contexts. They also ensure that the database remains applicable within its niche. For each iteration or instantiation of developing a linguistic content database, the context and the localised settings define the linguistic content and make-up of the database. Through this development work, the research questions and development work as outlined in Chapter 1 could be examined and progressed.

4.1.1 Local resource challenges: The Irish primary school classroom

In narrowing into the local research area of the Irish primary school classroom, the resource challenges associated with the local context are summarised below. Research has shown that there is a lack of resources (textbooks, teaching materials, teacher time, funding) reported at primary level, and that these are required to effectively teach the *Primary School Curriculum* (PSC) (Government of Ireland, 1999) (NCCA, 2005a, 2008). The emerging challenges and benefits of textbooks as a core resource in teaching and learning in Ireland have been drawn together in Table 4.1 and Table 4.2 below. These issues permeated the discussion in Section 2.2, and indeed overlap with the global issues identified above. Table 4.1 shows the benefits and challenges of textbook use from the teachers' perspectives and Table 4.2 shows the Inspectorate of the Department of Education and Skills' perspective surrounding the over-use of textbooks.

The varying perspectives of the teachers and DES Inspectorate across Table 4.1 and Table 4.2 are evident, where the DES Inspectorate were reported to perceive *any* textbook over-use or over-reliance as a challenge to and cause for concern in teaching and learning. The DES Inspectorate maintain that "textbooks should be used as resource material to support the implementation of the curriculum, rather than the basis of the planning for curriculum delivery" (DES, 2005a, p. 38). Teachers on the other hand, reported the presence of textbooks in their classrooms as having both benefits and challenges. The reported benefits reflected areas where teachers used textbooks as a scaffold for their teaching and as a substitution for the more established practices of school and classroom planning and referring to the PSC. Their cited

[11] The phrase 'modern language curriculum documents' will be used from this point forward to refer to the official curriculum statements and documents for the MLPSI.

challenges mirrored those reported by the DES Inspectorate with respect to planning and relevance to the PSC, but also added their contribution to curriculum-overload (as an additional text source) and parental expectation for completion of textbooks.

While the benefits and challenges associated with textbook use in Table 2.2 (p. 29) refer to local and global coursebooks, many of the textbooks in Irish primary schools are local textbooks, produced for the local Irish market and somewhat aligned to the PSC. Similar themes do emerge across the global and local Irish domains, the majority of which centre on the themes 'Basis in the local curriculum / Methodology / Examination Process'. The levels of localisation required for Irish textbooks are related less to cultural relevance as in the global instance, and more to their relevance to curriculum-based planning, teaching and learning, as well as the individual children in Irish classrooms.

<table>
<tr><th colspan="2">Teachers' view of textbook use</th><th rowspan="2">Potential solution and benefit</th></tr>
<tr><th>Benefits</th><th>Challenges</th></tr>
<tr><td rowspan="6">▪ Guide curriculum implementation
▪ Ensure standardisation across class levels (e.g. reading ability)
▪ Support classroom and whole-school planning
▪ Support teachers who may lack subject-knowledge
▪ Support teachers who find the curriculum documents difficult to navigate</td><td>Over-reliance for planning
Reduced use of the variety of methodologies outlined in the PSC for all subjects</td><td>Plan according to curriculum documents then source relevant resources. Ensure resources are curriculum-based
***Localised settings*:** Modern language curriculum documents and PSC</td></tr>
<tr><td>Impediment to cross-curricular or integrated learning</td><td>***Localised settings*:** Relevant areas of linkage and integration from the PSC</td></tr>
<tr><td>Parental expectation for textbooks to be completed</td><td rowspan="2">Replace textbook with teacher-generated resources</td></tr>
<tr><td>Adds to perception of curriculum overload – more books to complete or address</td></tr>
<tr><td>Need to be aligned to the PSC</td><td>Plan according to curriculum documents then source relevant resources. Ensure resources are curriculum-based
***Localised settings*:** Modern language curriculum documents and PSC</td></tr>
<tr><td>Unattractive, outdated, poor stimulus for learning, uninteresting for children, not relevant to children, lack of authentic texts</td><td>***Make resources suitable for local/individual teaching and learning settings –***
allow the teacher easy access to a source of relevant and tailored (cultural, level) materials.
Ensure resources are gathered and compiled to meet local teaching and learning needs</td></tr>
</table>

Table 4.1 Summary of benefits and challenges associated with textbook (over)-use from the viewpoint of teachers (benefits and challenges compiled from Gilleece et al, 2012; Harbison, 2008; NCCA, 2005a, 2008a; Varley, Murphy and Veale, 2008)

DES Inspectorate showed concern surrounding textbook (over)-use where they...	Perceived/Potential Impact	Potential solution and benefit
lacked **relevance or suitability** to the subject, curriculum or children being taught	Children do not learn the relevant content of the curriculum	Plan according to curriculum documents then source relevant resources. Ensure resources are curriculum-based ***Localised settings*:** Modern language curriculum documents and PSC ***Make resources suitable for local/individual teaching and learning settings –*** allow the teacher easy access to a source of relevant and tailored (cultural, level) materials. ***Ensure resources are gathered and compiled to meet local teaching and learning needs***
	Children are not provided with opportunities to fully engage with the curriculum and the range of methodologies contained in it	
	Teachers have difficulty in gauging what has been learnt and needs to be learnt across each level as it is not standardised according to the curriculum.	
	Teachers do not consider other important elements of learning such as prior learning, assessment or progression	
reduced the use of **authentic and varied** sources of texts	Teachers do not invest in their classroom's print-rich environment (e.g. authentic texts, posters, poetry books, novels) as they would if they were using a variety of sources of text for teaching and learning	System issue
	Children have reduced opportunities to engage with a variety of genres of text.	
	Children have reduced exposure to a variety of genres of text	
were used as a basis for **teacher and whole-school planning**, rather than the relevant curriculum documents	Teachers are less likely to teach particular strands or strand units (content) if they are not present in textbooks or programmes.	Plan according to curriculum documents then source relevant resources. Ensure resources are curriculum-based ***Localised settings*:** Modern language curriculum documents and PSC
	Teachers neglect / do not consider items of teaching and learning in the curriculum which are not included in textbooks (e.g. methodologies, assessment, progression, learning objectives).	
used/outlined **teaching methodologies** which were at odds with the teaching methodologies espoused in the curriculum or indeed offered impoverished opportunities to use all relevant and necessary teaching methodologies	Teachers overuse methodologies such as reading, writing and teacher questioning at the expense of other teaching methodologies (which vary across the subjects of the PSC).	***Localised settings*:** Modern language curriculum documents and PSC ***Make resources suitable for local/individual teaching and learning settings –*** allow the teacher easy access to a source of relevant and tailored (cultural, level) materials. ***Ensure resources are gathered and compiled to meet local teaching and learning needs***
	Children do not learn the relevant skills they should through engaging with the PSC.	
	Children's learning is focused on a limited range of methodologies which can lead boredom and a dislike of learning activities over time.	

Table 4.2 Summary of challenges associated with textbook (over)-use from the viewpoint of the DES Inspectorate and perceived/potential impacts of these challenges (potential impact compiled from DES, 2005, 2005a, 2007, 2009)

The potential solutions and benefits outlined in each case above, map the issues surrounding resources used in Irish schools to the requirement for curriculum-based resources, which are relevant to the children in Irish classrooms and empower the teacher to easily select and use relevant materials. ***Localised settings*** account for the specific boundaries or parameters of the research context of the Irish primary modern language classroom which include (i) specific national modern language curriculum documents, (ii) the PSC and (iii) links which can be made between them (among other parameters which emerge below in relation to language in the primary classroom). Localised settings will be described in Section 4.1.3 below in relation localising a database solution for the research context.

4.1.2 Local educational challenges: The complex language situation in Ireland

The English and Gaeilge/Irish language subject curriculums which form part of the Primary School Curriculum (PSC) (Government of Ireland, 1999) are generally taught in a disparate way; they are "considered and treated separately" (Council of Europe and DES, 2008, p. 32). Recent work from the NCCA (NCCA, 2014; Ó'Duibhir and Cummins, 2012) on revising and redeveloping a 'language curriculum' has focused on bridging the gap between the previously disparate language subjects (NCCA, 2014). A closer examination of the teaching and learning of Irish revealed that students' perceptions of the usefulness of learning Irish decreases as they progress through the education system (Smyth, Dunne, McCoy and Darmody, 2006). This could be due to the more traditional stance associated with how Irish is taught which can be far removed from how modern languages are taught and perceived.

Ireland's newcomer migrant population add another linguistic element to the English and Gaeilge/Irish language environment – English as an Additional Language (EAL) and the addition of newcomer languages to schools. Figures from 2007 estimated that 7.5% of children in primary schools were learning English as an additional language (Smyth, Darmody, McGinnity & Byrne, 2009) and that over 167 languages are now spoken in Ireland (NCCA, 2008). According to the 2011 Irish census data, languages now spoken in Ireland include majority as well as minority languages, such as Polish, Chinese, Urdu, Yoruba and Arabic (CSO, 2011) and illustrate the many continents from which migrants have travelled.

In tandem with the established PSC, some 546 schools incorporated a modern language up to the end of the school year 2011/2012 through their participation in the MLPSI. Schools were allowed to participate based on their suitability in maintaining a diversity of languages across Initiative schools, their availability of teaching staff to teach the modern language, the available MLPSI budget to support and fund additional schools in any one year and any other criteria of

focus within the MLPSI for any given year. Following the abolition of the MLPSI in the Irish 2012 budget (Department of Finance, 2012), it is hoped that as many of these schools as possible, will continue to offer a modern language to their fifth and sixth class children in whatever shape or form it is feasible.

Table 4.3 below, outlines the complexity of the language situation in schools in Ireland presently. Three main school-types emerge in Ireland, where the differentiating factor is the language of instruction[12]. (1) English-medium schools are schools where English is the language of instruction, (2) Gaeltacht schools are those where Irish is the medium of instruction and Irish is the home language of the majority of students and (3) Gaelscoileanna or all-Irish schools are schools where Irish is the language of instruction but not the home language of the majority of students. The second of the three school types is found exclusively in Gaeltacht areas of Ireland, where Irish still remains as the first language of the inhabitants. Table 4.3 illustrates the status of each of the languages English, Irish and modern languages (whether used/taught as the medium of instruction or as a subject) across the school types, and the presence of EAL in English-medium schools (English: Additional language – medium of instruction).

	English			**Irish**			**Modern language**
	L1	L2 – subject only	Additional language – medium of instruction	L1	L2 – subject only	L2 – medium of instruction	L2 – subject only
English-medium school	•		•		•		•
Gaeltacht school	•	•		•		•	•
All-Irish school	•	•		•		•	•

Table 4.3 Complexity of the language situation across school types in Ireland (Ó'Duibhir and Cummins, 2012, p. 27).

Aside from the complex language situation now present in Irish primary schools, challenges have emerged in relation to the teaching and implementation of English and Gaeilge/Irish at

[12] Other 'types' of schools are also apparent in Ireland, where the differentiating factor could be the religious ethos of the school or its public/private/independent status. Similarly, there are independent schools where German or French, for example, are the languages of instruction.

primary level (NCCA, 2005a, 2008a; DES, 2005, 2005a, 2007). Furthermore, Ireland's literacy and numeracy levels were reported to be lower than the European average in the PISA 2009 study (Perkins, Moran, Cosgrove & Shiel, 2010) which focused on literacy in this cycle. PISA 2012 results saw literacy levels gaining strength from the 2009 dip (Perkins, Shiel, Merriman, Cosgrove & Moran, 2013). Rather than juggling existing curriculum time to reallocate more resources to literacy and numeracy as suggested by the DES (DES, 2010), the NCCA suggested a more thorough solution – to revise the curriculum rather than juggle the overloaded existing one which teachers need to currently "divide… (time) and conquer… (all subjects)" (NCCA, 2010b, p. 28). The current DES strategy (2011-2020) (DES, 2011) to overcome challenges with literacy and numeracy include defining learning outcomes for the English and Irish Curriculums, as well as a potential reallocation of curriculum time to allow for greater coverage of literacy and numeracy development.

The NCCA are currently working on revising the language curriculum, to one which will take account of the complex linguistic school environment of today, incorporate the language area of *Aistear: the Framework for Early Learning* (NCCA, 2009) and teach languages across the curriculum. This revised language curriculum will "move away from teaching language and literacy in discrete subjects and towards providing more integrated learning experiences for children. The new specification for the language curriculum will highlight connections across learning experiences at three levels of integration" (NCCA, 2012, p. 60). This integrated language curriculum was to be achieved through the recommendations made by Ó'Duibhir and Cummins (2012):

- "explicitly draw children's attention to similarities and differences between their languages [Language Awareness]
- pay more attention to the role of literacy in supporting language development in L2 contexts in particular.
- include more explicit differentiation for the needs of L1 and L2 learners
- devise one L1 curriculum which children would follow in English or Irish depending on school context and child background. The L1 curriculum should be cross-referenced with the L2 and modern language curricula and use largely the same structures and descriptors" (p. 16).

The new draft language curriculum spans early years education (junior infants to second class, ages 4/5-8/9) (NCCA, 2014). It is "an integrated curriculum—it has the same curriculum structure and components for English and Gaeilge to support integration across the two languages" (NCCA, 2014, p. 6). It also provides a clearer mapping of progress and milestones that children should achieve in their learning across these years through a "continuum (map)" (NCCA, 2014, p. 6). This latter feature was missing from earlier language curriculum

documents, where a play and fun-based communicative curriculum at primary level for Irish, for example, (Government of Ireland, 1999) mismatched the more text- and form-focused ethos of the post-primary curriculum for Irish. There was some confusion surrounding systematic progression at primary level (NCCA, 2005a), and at what stage children should be learning the different features of the languages. This affirms the argument made by Decoo (2011) in which he outlined evidence from the 1970s to present day of the importance of a "curricular continuum" (p. 2) and pedagogical "sequencing" rather than the use of "criterion … to satisfy on the spot the language needs elicited by a topic or a setting" (p. 3, citing Girard, Courtillon, Page and Richterich, 1988). Furthermore, Decoo (2011) cited Byrnes' (2001) definition of curriculum as "the attempt to devise a sequence of educational opportunities for learners that builds on internal interrelations and continuities among the major units of instruction … to enhance learning. Critical considerations are the selection of content and its sequencing…" (p. 25). The NCCA have recently completed a consultation process on the draft and plan further engagement with schools into the 2014/15 and 2015/16 school years to progress the work (NCCA, 2014).

Existing classroom resources for teaching English, Irish and modern languages mirror the individual and disparate nature of the national curriculum documents they support. With integration, linkage and an integrated language education at the core of ongoing language curriculum developments, classroom resources are required to follow the more cohesive nature of the curriculum revisions. With national curriculum change underway, a more flexible and reusable approach to resource generation would optimise the time and effort spent in their creation, while allowing customisation to the changing curriculum while under revision and in its final state. Table 4.4 below outlines the contextual features and the related resourcing requirements.

Context features	Perceived/potential impact	Potential solution and benefit
Language Curriculums treated in a disparate way	Resources for languages mirror disparate nature of curriculums	Make **links** between all language curriculums and language learning **Localised setting**: PSC, modern language curriculum documents, Curriculum Integration and Linkage, Language Awareness
Many languages as L1/L2 in the Irish classroom (Irish, English, EAL and/or modern language)	Rich linguistic environment in schools which is not drawn upon to enhance overall language learning	Make **links** between all language curriculums and language learning **Localised setting*:*** Language Awareness, Learner Error Analysis (given specific L1 and L2)
Concerns about national literacy levels	Challenges relating to reading and writing, and maximising curriculum time for literacy development	**Curriculum integration** to create links between subjects and save individual subject time. **Language Awareness to assist in language learning skills**. **Localised setting**: PSC, modern language curriculum documents, Curriculum Integration and Linkage, Language Awareness
National Curriculum under revision	Resources will be required for the revisions under trial in schools and the revised curriculum once finalised	**Reusable and customisable resource generation** which can be tuned to the relevant shape of the changing curriculum.

Table 4.4 Educational features of the research context and associated resourcing requirements

Three additional localised settings emerged from the challenges surrounding languages in Irish primary schools:

(i) Curriculum integration and linkage

(ii) Language awareness

(iii) Learner Error Analysis (L1=English, L2=Irish, L3=German)

4.1.3 Towards a local solution to resourcing and educational challenges

The local context along with its associated resourcing challenges and language diversity, all feed into defining the "contextual requirements and parameters" research input of the problem space outlined in Figure 3.11 (p. 105). A closer examination of this research input illustrates that it is this input which defines the particular content of any one instantiation of a linguistic content database. Figure 4.1 below zooms into this research input to illustrate the contextual impact on the database content. The linguistic content database, in turn, comprises curriculum-based raw linguistic content and localised setting content.

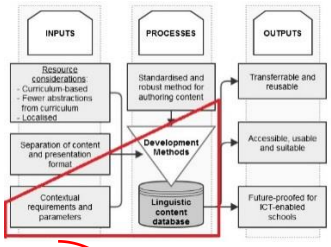


Figure 3.11 (p. 105)

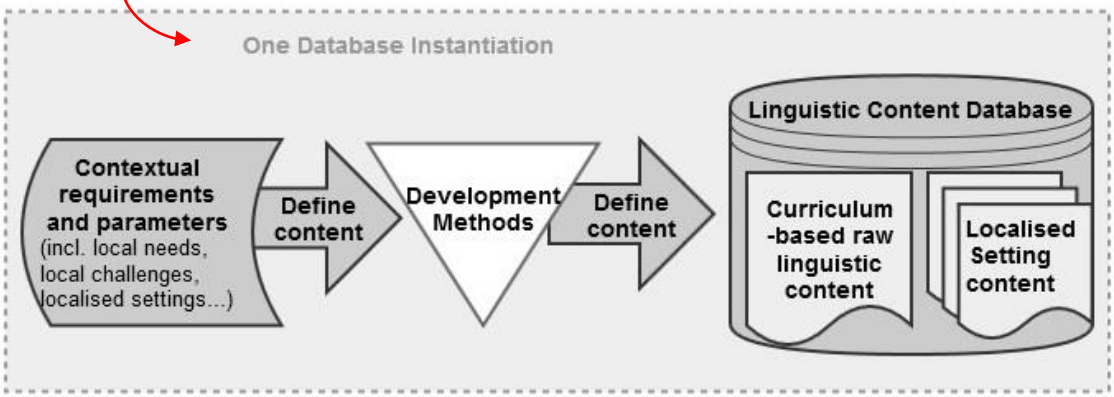


Figure 4.1 Defining localised database content for one database instantiation

The local resource challenges outlined in Section 4.1.1 echo those global ones identified in the preceding chapter and are accounted for in the overall problem space in Figure 3.11 (p. 105). The more specific local resource challenges as well as challenges relating to language presence, teaching and learning in Ireland are accounted for in their inclusion in the "contextual requirements and parameters" research input. Figure 4.2 below illustrates the instantiation of the diagram for the local Irish primary school context and its associated challenges.

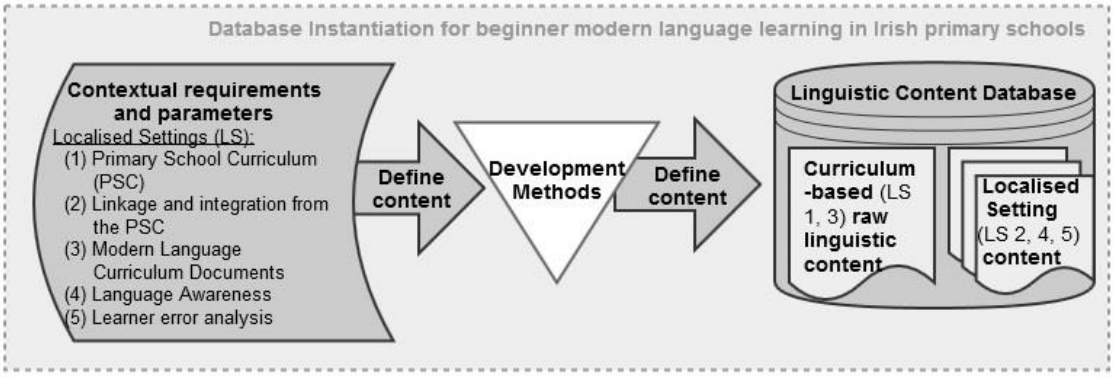


Figure 4.2 Database instantiation for the research context

The development methods employed to initially author, collate and test the raw linguistic content are outlined below from Section 4.2 onwards. The raw linguistic content testing as well as data gathering on the localised settings took place in a primary school classroom in Dublin, Ireland. The target modern language used was German. In acting as a visiting German teacher and

researcher in the classroom, I was able to gather the relevant data according to the methods outlined in Section 4.2 below.

Localised settings allow the database solution to be specific to the primary Irish classroom whilst the selection or omission of localised settings allow the database to continue to be reusable in other contexts. Table 4.5 below illustrates how localised settings can be switched on or off, visible or hidden, to enable relevant contextual content to be selected in one or many settings. The first column illustrates how the localised settings set out for this research are all visible/relevant to the database. The second column illustrates how these same localised settings can apply to a similar context, that of another primary German language learning context, in which case the localised settings relevant to the specific Irish curriculum are hidden. The final column indicates how a primary French language learning context can draw on German content as the basis of an equivalent course in French. The specific background for the localised settings of the research context of the Irish primary school modern language classroom are explored in Section 4.1.4 below.

Context	Irish primary modern language classroom (German)	Beginner primary learners of German (CEFRL level A1)	Beginner primary learners of French (CEFRL level A1)
Localised Settings	■ National modern language curriculum documents ■ German linguistic content ■ PSC ■ Curriculum links	■ National modern language curriculum documents ■ German linguistic content □ PSC □ Curriculum links	■ National modern language curriculum documents □ German linguistic content ■□ French equivalent of German linguistic content □ PSC □ Curriculum links

Table 4.5 Illustration of how localised settings account for the reusability or specificity of database content (Note: ■ Visible/relevant database content; □ Hidden/irrelevant database content; ■□ new additional localised setting required which builds on existing content)

4.1.4 Background information for localised settings

Five localised settings were identified for this research context:

(1) the PSC,

(2) linkage and integration topics from the PSC

(3) the modern language curriculum documents

(4) Language Awareness for Gaeilge/Irish, English and German and

(5) Learner error analysis.

The context and background for each localised setting are described below.

(1) The Primary School Curriculum (PSC)

The Primary School Curriculum (PSC) (Government of Ireland, 1999) forms the foundation for teaching and learning across its six curriculum areas and eleven subjects. Additional subjects, like German, cannot be taught without due consideration being given to other learning taking place in the same context and the potential transfer of learning from one subject to another (integration) or within subjects (linkage). Similarly, the PSC is a well-defined curriculum which uses a common format across its subjects. Knowledge of its structure and content are essential in creating the opportunities for linked and integrated learning.

In order for the raw linguistic content to be curriculum-based, an examination of the Primary School Curriculum (Government of Ireland, 1999) was necessary – its structure and contents – as well as the existing languages within the established Primary School Curriculum (English and Gaeilge/Irish). An examination of the PSC was also relevant in acting as an effective practitioner in the primary school setting.

Curriculum Area	Subject
Social, Environmental and Scientific Education (SESE)	History
	Geography
	Science
Language	Gaeilge / Irish
	English
Arts Education	Visual Arts
	Drama
	Music
Mathematics	Mathematics
Social, Personal and Health Education	Social, Personal and Health Education
Physical Education	Physical Education
Religious Education	Religious Education

Table 4.6 The Primary School Curriculum

The revised Primary School Curriculum (Government of Ireland, 1999) replaced Curaclam na Bunscoile (1971). The revised curriculum comprises 11 subjects divided up into six curriculum areas (Table 4.6). Religious education represents an optional 12th subject and adds a seventh curriculum area. It is determined by the relevant church authority.

The Introduction to the Primary School Curriculum also makes reference to the contribution that a third language can make to language acquisition at primary level "… in the context of the outcome of the pilot project on modern languages [later renamed the MLPSI]" (Government of Ireland: Introduction, 1999, p. 43).

Each subject's curriculum comprises an "introduction, aims, broad objectives, overview tables, content, concept and skills development, guidance on the selection of content [and] assessment statement" (Primary School Curriculum: Introduction, 1999, p. 41). The content is further broken down into strands, strand units and content objectives. Figure 4.3 below outlines how the structure of the content is hierarchical, the strands break down into strand units, which in turn break down into detailed content objectives. The curriculum content is repeated over each of the four primary school levels.

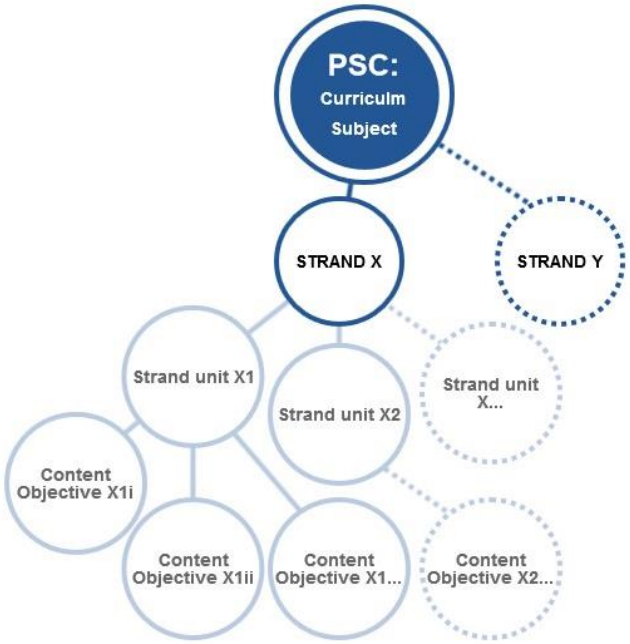


Figure 4.3 The hierarchical nature of the contents of the PSC

Theories of learning underpinning the PSC

The revised PSC was published in 1999, with child-centred theories underpinning the approaches and methodologies outlined within it. The revised curriculum inherited the pedagogical principles of its predecessor, Curaclam na Bunscoile (1971), and added a few additional principles which took account of the "educational, economic, social and cultural developments in Irish society" (Government of Ireland, Introduction, 1999, p. 2) which had changed in the interim.

Three broad pedagogical principles form the foundation for the principles of the PSC (Government of Ireland, Introduction, 1999, p. 8):

- "Activity and discovery methods
- An integrated curriculum
- Environmental-based learning."

The principles underlying the primary curriculum in Ireland are based on different theories of child development and growth, including the theories of Piaget, Bruner and Vygotsky on how children think and learn (INTO, 2008, p. 5). The PSC on a whole is based on constructivist

(Vygotsky and Bruner) and Piagetian theory. While an in-depth discussion of constructivist theory is beyond the scope of this research, in brief, constructivism places importance on meaningful, authentic activities that help the learner to develop problem solving skills and construct their own understanding (Wilson, 1996). Vygotsky adds the importance of the social element to cognitive development and maintains that social interaction and social context are required for learning to take place (Vygotsky, 1978). Bruner places emphasis on learning being an active process in which learners construct knowledge based on their existing and past knowledge. Bruner devised a method which described how people "modify, expand and adapt existing concepts to meet new demands" (Child, 1986, p. 160). "Piaget takes an 'active' rather than a 'passive' view of the part played by children in their own development. The child's self-directed motor activity is seen as a necessity in cognitive development" (Child, 1986, p. 146).

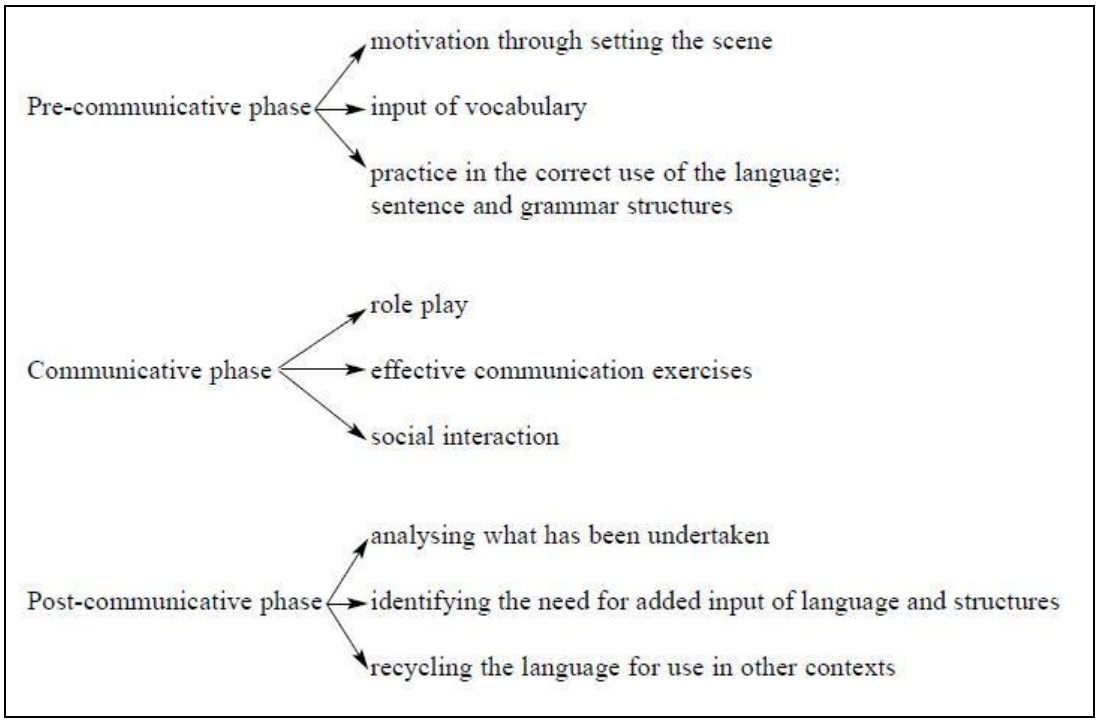


Figure 4.4 The three phases of the communicative approach (NCCA, 1999, p. 7)

A communicative approach to language learning places the emphasis on learning language necessary to communicate in the target language rather than being able to respond to fixed phrases. Figure 4.4 outlines how a communicative approach is divided into three phases in the classroom – (1) Pre-communicative phase, (2) Communicative phase and (3) Post-Communicative phase. The pre-communicative phase provides children with the necessary language (words and phrases) for the communicative task ahead. This phase involves the revision of relevant and previously known language and the introduction of new language. The communicative phase provides children with the opportunity to practise the language they know and have learnt in the previous phase in a communicative activity; that is, an activity where they are actively using the language (e.g. through role-play activities). The post-communicative phase

helps to reinforce language learnt, assesses progress made and language produced, and sets the scene for the next iteration of the cycle.

A task-based learning approach complements a communicative approach in that it involves the authentic use of the target language to complete tasks such as asking for directions. In the context of the Irish Curriculum for example, these tasks are based around the language topics/themes (Téamaí) (see below).

Language in the Curriculum and the Language Area

Language and references to language permeate the revised curriculum; language is the means through which we think and access information. One of the principles of the revised curriculum is that "language is central in the learning process" (Primary School Curriculum: Introduction, 1999, p. 8). One of the principles of learning is "learning through language", where language is cited as having a crucial role to play in the development of children (Primary School Curriculum: Introduction, 1999, p. 15).

Key issues in education outlined in the revised curriculum include *the Irish language* as a basis for Irish children learning two languages (along with English) from the first year of their education and as a basis for extending each child's linguistic experience and deepening cultural awareness (Primary School Curriculum: Introduction, 1999, p. 27). A second key issue relating to language is "the European and global dimension" which refers to the pilot project on modern languages in primary schools (later named the MLPSI) as a basis for enhancing children's linguistic and cultural awareness, as well as realise their individual, national and European identity (Primary School Curriculum: Introduction, 1999, p. 27).

Two specific aims of the curriculum refer to language, (a) "to enable children to develop literacy skills, comprehension skills and expressive skills in language and to appreciate the power and beauty of language" and (b) "to enable children to develop a respect for cultural difference…" (Primary School Curriculum: Introduction, 1999, p. 34). Nine of the twenty-five general objectives outlined in the curriculum pertain to language (Primary School Curriculum: Introduction, 1999).

The Primary School Curriculum suggests a "minimum weekly time framework" (Primary School Curriculum: Introduction, 1999, p. 70) in which school time is allocated to curriculum areas and subjects. The framework makes a distinction between infant classes and more senior classes. Infant classes have a shorter school day, typically finishing about one hour earlier than more senior classes. It is suggested that approximately *one third* (30%) of overall curriculum time is designated to the language area with the ratio of 8:7 (L1:L2) representing the time to be

attributed to the children's first language and the children's second language. This time allocation reinforces the importance that language is allocated in the Primary School Curriculum.

In comparison to our European counterparts, Ireland devotes the greatest amount of curriculum time (30%) to *reading, writing and literature* in L1 and L2 (alongside France and Mexico) (OECD, 2012)[13]. When we add in the time OECD countries devoted to modern languages, Ireland remains above the OECD country average of 29%[14] of curriculum time for *reading, writing and literature* and *modern languages* (calculated from OECD, 2012). Any additional rebalancing of curriculum time towards literacy, as suggested by the DES (2011) to overcome national literacy challenges, will place a more pronounced weight on the language/literacy curriculum balance, and add to the supporting resources required.

Language and cultural awareness are elements which permeate the entire PSC, as well as having a more dominant presence in the language area. Language awareness is examined below within its own localised setting. Cultural awareness within the PSC features in the "key issues in primary education" within the contexts of the "Irish language" and the "European and global dimensions" (Government of Ireland: Introduction, 1999, p. 27) and focuses on developing children's awareness of their national heritage as well as their citizenship in a globally diverse community. Specific aims and general objectives of the PSC also account for children being enabled "to develop a respect for cultural difference" (Government of Ireland: Introduction, 1999, p. 34). The PSC also merits learning a second language within the context of learning Irish or English as a second language, as adding "a further dimension to the child's sense of language and cultural awareness" (Government of Ireland: Introduction, 1999, p. 43). Cultural awareness also features in the Introduction to the Drama Curriculum and the Music Curriculum.

The English Curriculum

The English Curriculum does not differentiate for schools where English or Irish is the language of instruction (as for the Irish Curriculum, see below). A restructuring and representation of the English Curriculum was carried out in 2005 to flip the strands and strand units of the curriculum to make them more accessible and usable for teachers (NCCA, 2005a). The curriculum now has three strands, (1) oral language, (2) reading and (3) writing. There are four strand units, (1) receptiveness to language, (2) competence and confidence in using language, (3) developing cognitive abilities through language and (4) emotional and imaginative development through

[13] Examining the OECD data for 9-11 year olds, as the next bracket of 12-14 year olds would incorporate secondary school teaching/curriculums in Ireland.
[14] OECD countries, such as USA and Switzerland where allocation of curriculum hours differ by federal state/canton, and which did not have number entries, were excluded from the averaging calculation.

language. The English Curriculum places great importance on oral language and literacy skills. A further redevelopment of the English Curriculum was developed in 2010 which represented the content in a tabular format and divided the content objectives into two further categories – *development* and *skills*. The many guises the English curriculum has taken since 1999 reflect the difficulties teachers have had with accessing and using its contents for planning and teaching.

The Irish Curriculum / Curaclam na Gaeilge

The Primary School Curriculum refers to *Gaeilge* as the subject rather than *Irish*. Irish is taught in three different types of school as outlined in Section 4.1.2 above. There are two Irish curriculums to match the two different mediums of instruction of the three school types; one for English-medium schools and one for Irish-medium schools. The main difference between the two curriculums is the period when Irish reading and writing are introduced. The school in which the classroom research was carried out was an English-medium school where Irish was taught as a subject.

The four strands of the Irish Curriculum are (1) Listening / Éisteacht, (2) Speaking / Labhairt, (3) Reading / Léitheoireacht and (4) Writing / Scríbhneoireacht. The curriculum promotes learning language topics/themes (see Table 4.7 below) which are relevant to children in their everyday lives. It also promotes a communicative approach and making Irish learning enjoyable through the use of active learning activities for children including songs, stories, rhymes and riddles (NCCA, 2008a). This approach matches the approach outlined in the *Pilot Project on Modern Languages in the Primary School Draft Curriculum Guidelines* (NCCA, 1999) (see below) which allowed greater links to be created between German and Irish content for the linguistic content database and classroom teaching.

Myself / Mé féin	Shopping / Siopadóireacht
At home / Sa bhaile	Pastimes / Caitheamh aimsire
School / An scoil	Clothes / Éadaí
Food / Bia	The weather / An aimsir
Television / An teilifís	Special occasions / Ócáidí speisialta

Table 4.7 Curaclam na Gaeilge: Themes / Na téamaí

(2) Linkage and integration topics from the PSC

The Primary School Curriculum (PSC) works as a whole in incorporating intra-subject and cross-curricular knowledge and teaching through *linkage* and *integration*. Linkage and integration form an essential component of the PSC for teachers, in allowing their teaching to cross between separate sections of the same subject (linkage), or cross between different subjects (integration). The PSC refers to linkage and integration as both being types of integration:

- *Linkage*: integration within a subject
- *Integration*: Cross-curricular connections.

Linkage and integration were vital to the development of the database of raw linguistic content with localised settings in so far as they provided the relevant connections between the modern language (German) and the rest of the PSC, ensuring that connections made were not contrived, rather based on the established content of the PSC. In the case of Gaeilge/Irish, the ten themes outlined closely match those outlined in the modern languages curriculum documents, and these represent a mere starting point for the connections which can be drawn between the modern language and the PSC as a whole.

The inclusion of integration in the PSC means that it is organised so that it cuts across subject–matter lines, bringing together various aspects of the curriculum into meaningful association to focus on broad areas of study (Shoemaker, 1989). The integrated nature of the PSC is one of the underpinning principles of the PSC, and indeed, the basis of one of the three pedagogical principles of the PSC. The PSC endorses that "learning is most effective when it is integrated" (Government of Ireland: Introduction, 1999, p. 9). The PSC also "recognises the integrated nature of knowledge and thought and stresses the connections in content in the different curriculum areas. This creates harmony in the child's learning experiences and serves the complex nature of the learning process" (Government of Ireland, 1999, p. 11). The PSC states younger children should be less aware of subject boundaries in what they learn through integration and older children should become aware of the "interconnectedness of knowledge and ideas" through integration in the learning process (Government of Ireland, 1999, p. 16). References to areas of linkage and integration are included in the Curriculum Statements of the subjects of the PSC.

More recently, Ó'Duibhir and Cummins (2012) drew further conclusions on integration with regard to the PSC and language - they distinguished three dimensions of integration that apply to language curricula: "(i) integration within the teaching of a specific language [linkage], (ii) integration across the curriculum, and (iii) integration across languages" (Ó'Duibhir and Cummins, 2012, p. 27/28).

Beyond the realm of the PSC, many sources discuss the various terms which can be used to encompass the concept of integration (e.g. theme-based teaching, topic-based teaching, problem-based teaching, project-based teaching, unit-based teaching, interdisciplinary teaching or synergistic teaching) and they originate from the many domains of education. These sources also posit that there are degrees of integration (e.g. Fogarty, 1991; Palmer, 1991) and outline the rationale and the advantages associated with incorporating integration across curriculums (e.g.

Oster, 1993; Shoemaker, 1989). Suffice to say, that all sources referring to integration are unanimous in their declaration of the advantages associated with a connection being "discerned between skills and content that transcends curriculum lines" (Lake, 1994), no matter what form it takes or what it is labelled. For the purposes of this research, linkage and integration are confined to the definitions and uses outlined in the PSC.

(3) The Modern Language Curriculum Documents

The modern language curriculum documents (NCCA, 1999, 2001) formed the starting point for the development of the raw linguistic content. Their structure and format aligned to the PSC format. They also represented the core approved content for modern language teaching in Irish primary schools. The curriculum documents which were drafted for use on the pilot project on modern languages in primary schools (later renamed the Modern Languages in Primary Schools Initiative (MLPSI)) comprise the *Pilot Project on Modern Languages in the Primary School: Draft Curriculum Guidelines* (NCCA, 1999) and *Modern Languages in Primary Schools. Teacher Guidelines* (NCCA, 2001). This section describes the content and layout of the *Draft Curriculum Guidelines* (NCCA, 1999) with the relevance lying in ensuring the raw linguistic content is curriculum-based. The methodological underpinnings outlined in the associated *Teacher Guidelines* (NCCA, 2001) are also described.

The draft curriculum guidelines for modern languages in primary schools were devised on the basis of a two year teaching programme for the upper end of primary school in 5th and 6th classes (children aged 10-12). The MLPSI recommended that 2 x 45 minute language lessons per week formed the basis of modern language teaching in primary schools.

As with all subjects in the Primary School Curriculum (PSC) in Ireland, the curriculum statement for any one subject comprises the Curriculum document and the Teacher Guidelines document. The *Draft Curriculum Guidelines* were developed at the same time as the revised PSC was being developed, and so sits under the same teaching ethos which is espoused in the PSC. The associated *Teacher Guidelines* came at a later stage, but also sit within the context of the PSC.

In keeping with the layout and aligning somewhat to the content of the English and Irish Curriculums, the *Draft Curriculum Guidelines* set out 3 strands and 4 strand units. The 3 strands of the curriculum are (1) Communicative competence, (2) Language awareness and (3) Cultural awareness. The four strand units reflect the four language skills, (1) Listening, (2) Speaking, (3) Reading and (4) Writing. The curriculum notes that while the strands and strand units are set out separately, they should be used and developed in coordination with one another, as an integrated curriculum. Language functions and 10 language topics around which the strands and strand units can centre are introduced.

The strand communicative competence is described in terms of the communicative approach to learning (NCCA, 2001) and is divided into the three phases of communication as outlined in the previous section.

Language Awareness (LA) will be described below as a localised setting. In the context of the Draft Curriculum Guidelines, the definition of LA is more in keeping with its established use in language learning, than it is for either the English Curriculum or Curaclam na Gaeilge (Irish language Curriculum). The curriculum makes reference to children using other languages in their environment – English and Irish – to compare and contrast against the modern language being learnt. It is suggested that LA can be included by having "children focusing on the alphabet, sentence structure, conventions of language, the language appropriate to particular situations and simple rules applicable in the languages" (NCCA, 1999, p. 8). The curriculum focuses on the similarities between words in known languages and those in the modern language, as well as differences, such as the lack of noun gender in English.

The description of the strand **Cultural Awareness** draws attention to children learning about things that are relevant to them in the target language countries, for example, "homes, schools, holidays, food, feast days, customs and traditions" (NCCA, 1999, p. 9). It also makes reference to children learning about cultural differences and norms, potentially through the study of stereotypes.

The curriculum also sets out ways in which the modern language can be integrated into the school environment (e.g. through the use of classroom language) and explored outside the classroom (e.g. through the use of ICT).

The **ten language topics**, through which the stands and strand units are as follows:

- "introducing oneself
- talk about family
- where I go to school
- classroom language
- numbers, days of the week, months, birthdays, festivals, seasons and weather
- time
- where I live
- housing
- feast days and holidays
- hobbies"

(NCCA, 1999, p. 44-49):.

Language functions (NCCA, 1999) are outlined on a language by language basis and provide the basis for meaningful communication. The language functions are centred on the following areas:

- "meeting and greeting people
- giving and seeking information, finding solutions
- making requests and suggestions
- indicating opinions and preferences."

(NCCA, 1999, p. 14)

The MLPSI staff produced a European Language Portfolio (Council of Europe, 2006) called MyELP, which is based on the *Draft Curriculum Guidelines*. ELPs have their basis in the Common European Framework of Reference for Languages (CEFRL) (Council of Europe, 2001). The CEFRL allows parallels to be drawn between levels of language learning reached in domains where they are used. It is a global framework, which sets "clear standards to be attained at successive stages of learning and for evaluating outcomes in an internationally comparable manner*"* (Council of Europe, 2001). The CEFRL sets out six levels of modern language proficiency – Basic (A1, A2), Proficient (B1, B2) and Independent (C1, C2) ranging from beginner (Breakthrough) to advanced learners (Mastery). The draft curriculum guidelines for teaching modern languages in primary schools have been mapped to level A1 (Basic User: Breakthrough) of the Common European Framework of Reference for Language (CEFRL) (Council of Europe, 2001) (with scope also provided in MyElp for learners to explore extending their language competence to level A2 (Basic User: Waystage)). This mapping to the CEFRL means that the raw linguistic content will be transferrable to any global English L1/German L2 context up to the level of A1 on the CEFRL, although some alterations may be required for an

older learner group (as the register for children will be lower, and additional vocabulary could also be required).

Teacher Guidelines (NCCA, 2001)

The *Modern Languages in Primary Schools Teacher Guidelines* (NCCA, 2001) reflect the layout of the other PSC teacher guidelines. They comprise background information on modern language learning and the role of the teacher for example, the content of the modern language curriculum, guidelines for school and classroom planning, details of approaches and methodologies and exemplar lessons for each of the four modern languages.

As the contents of the teacher guidelines mainly focus on school and teacher-based classroom planning and methodologies, they will not be discussed in too much detail in this research. These areas of teaching and learning are relevant to the teacher and school who plan the course of language progress and lessons, *prior* to sourcing relevant resources (consumable resources) for use in the classroom.

(4) Language Awareness for Gaeilge/Irish, English and German

The concept of Language Awareness has been attributed to Eric Hawkins who "proposed a new subject, 'language', to be taught as a 'bridging subject', linking English and the foreign language in the curriculum" (Hawkins, 1999, p. 124 citing Hawkins 1974). Language Awareness (LA) has been defined as a mental attribute which develops through paying motivated attention to language in use, and which enables language learners to gradually gain insights into how languages work. It is also a pedagogic approach that aims to help learners to gain such insights (Bolitho, Carter, Hughes, Ivanic, Masuhara & Tomlinson, 2003). As the name Language Awareness implies, LA tries "to guide learners towards discovery and understanding of truths about language" (Svalberg, 2001, p. 200) or provide the learner with a greater awareness of language.

Irish schools present an interesting linguistic setting. Most children in Ireland have experience with a minimum of two languages – English and Irish (in some cases an exemption can be applied for for Irish, e.g. where a child has a learning challenge like dyslexia, has studied outside of Ireland up to or beyond the age of 11 years or whose home language is neither English nor Irish). Evidence has shown that languages in Ireland are generally taught in a disparate way, without any links being drawn between languages being learned (Little, 2003; Council of Europe and DES, 2008). The publication of a Language Policy Profile for Ireland (Council of Europe & DES, 2008) has brought together many sources of evidence that mutually reinforce the lack of a uniform approach to teaching languages in the education system in Ireland. The profile refers to

the need for an integrated approach to language teaching which combats the current system of teaching languages "in isolation from one another" (Council of Europe & DES, 2008, p. 9). When we add the linguistic diversity that children for whom English is an additional language (EAL) or indeed traveller children with their own language (called Cant) bring to the classroom, the opportunities for making links between languages and drawing on children's existing knowledge of languages (Language Awareness) are interesting and many. Language awareness was included as one of the localised settings.

LA has been noted as a method to break the cycle of teaching languages in isolation from one another and begin to draw connections between them (Candelier, 2004; Hawkins, 1999). Where data has been gathered on the use of LA in Irish schools (NCCA, 2008, 2008a), evidence has suggested that few teachers cross the language curricula in their use of LA (NCCA, 2008, 2008a) and that there is a need for resources to support the cross-language use of LA (NCCA, 2008).

Language Awareness is explicitly referred to in the Irish Curriculum under the concept *feasacht teanga* and is referred to in the English Curriculum under the strand unit *receptiveness to language*. Neither the Irish nor the English reference to LA go far in describing what LA is nor in describing how teachers should incorporate it into their teaching. In fact, the only valid reference to and definition of LA in the Primary School Curriculum is found in the *Draft Curriculum Guidelines* for the Pilot Project on Modern Languages in the Primary School (NCCA, 1999). NCCA's more recent work on a Language Curriculum for early childhood and primary years (NCCA, 2014; Ó'Duibhir and Cummins, 2012) provides evidence of an educational move to take account of all languages in school environments in Ireland.

Through ensuring that language teaching resources take account of LA in all languages being taught in the primary school, a uniform approach to language teaching can commence. Perhaps what the Irish government and curricula are only starting to address through considering transferrable and universal language education in its system can in the shorter and longer term be realised and supported through the language teaching resources used in the classroom. The inclusion of LA as a localised setting was an attempt to bridge the gap between the time curriculum revisions will take to embed, and the more immediate local resource need for education across languages.

(5) Learner error analysis

> "Learners' sentences reflect their temporary language systems rather than their imperfect grasp of the target language. If a student makes a 'mistake',

> it is … an inevitable and natural part of the learning process" (Cook, 2012, p. 14-15).

Children in the participating classes already had knowledge of English (L1) and Irish (L2). When German (L3) was being added to their linguistic repertoire, it was an ideal opportunity to gather information on their typical learner errors. This information insured that typical and common errors for the participating children could be predicted, diagnosed and accounted for in their language learning through the resources being used in the classroom.

The process of learning a new language will inevitably require that the learner produces errors. In a traditional classroom setting, these errors have typically been observed by the teacher and addressed through a number of means, such as teacher-assisted self-correction, teacher correction or peer-correction (Walz, 1982). Language learners can make errors similar to those made by children learning their L1. These errors can be produced when a learner overgeneralises a rule to produce an incorrect form (Yule, 2010; Ellis, 1997), such as **goed* rather than *went*, or when learners make errors of omission, and for example, leave off the –s for plural nouns in English (Ellis, 1997).

Errors are sometimes associated with learners drawing knowledge of their L1 into the target language (TL). These errors are referred to as negative transfer errors (or interference). Knowledge from the L1 can also transfer successfully to the TL, in which case it is referred to as positive transfer. "Transfer is governed by learners' perceptions about what is transferrable [from their L1] and by their stage of development" (Ellis, 1997, p. 54). Negative transfer tends to be more common when learners are beginners in the TL and generally tend to decrease as the learner's competence in the TL increases (Yule, 2010; Ellis, 1997). Both positive and negative transfer errors are relevant to the research context as they help the children draw a conscious focus on what they already know about languages in their repertoire and what they already know in the target language, German. The inclusion of error analysis as a localised setting meant that the resulting linguistic content database could also assist teachers in pre-empting learner errors and drawing parallels through LA on typical positive and negative transfer errors among the children's language(s).

In acknowledging that transfer errors are not the only source of learner errors and in fact, have been reported to account for few of the errors made by learners (Ellis, 1997; Lightbown & Spada, 2006), the notion of interlanguage as well as differentiating between learner mistakes and errors come into play. Selinker (1972) first proposed the notion of an interlanguage to describe the "linguistic system*"* (p. 214) which is in place for learners as they navigate their way through a TL and which can account for errors made by the learner which do not stem from their L1 or the

TL. Cook 2012 summarised the relationship between a learner's L1, and the target language (TL) (be it L2, L3 or beyond) and the interlanguage, as below.

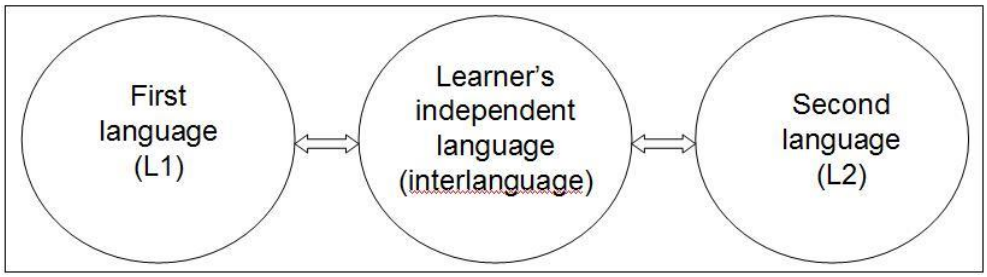


Figure 4.5 The learner's independent language (interlanguage) (Cook, 2012, p. 14)

Analysing a learner's interlanguage is a good basis for predicting what stage the learner has reached in their language learning, but also providing a basis for language teaching (Ellis, 1997). However, given that an interlanguage is an internal system, the examination of a learner's interlanguage involves "focus[ing] our analytical attention upon the only observable data to which we can relate theoretical predictions: the utterances which are produced when the learner attempts to say sentences of a TL" (Selinker, 1972, p. 213-4).

Corder (1974) and Ellis (1997) distinguished between learner errors and mistakes, where the former marks a shortcoming in the learner's knowledge of the TL and the latter is a 'slip' made by the learner, where he/she has previously demonstrated the correct use/form of a word or sentence. Thouësny (2011) referred to James' (1998) method for distinguishing between errors and mistakes, where the former cannot be corrected by the learner "without other sources of assistance" (Thouësny, 2011, p. 30) and the latter can. Thouësny (2011) provided an extensive overview of the differences between learner errors and mistakes and how to distinguish one from the other. Such a discussion is beyond the scope of this work.

Ellis (1997) presented four stages associated with analysing learner errors – (1) identify errors, (2) describe errors by classifying them according to types, (3) explain the origin of the errors (e.g. an interlanguage 'rule') and (4) evaluate the learner errors. The extent to which these four stages are relevant to any one research project depends on the aim of the project. Identifying, diagnosing (describing/explaining) and correcting (evaluating) learner errors are stages which are more relevant, for example, to computational systems (or to some Computer-Assisted Language Learning (CALL) systems) which are developed to assist language learners (Heift and Schulze, 2007), whereas identifying and correcting learner errors are stages which are more relevant to the language teacher in the classroom. The first two of these processes (identifying and diagnosing) were relevant to analysing learner errors for their integration into the linguistic

content database, while the latter two (identifying and correcting) were useful in acting as a visiting German teacher in the research classroom.

4.2 Research design overview

The aim of the research was to account for localisation in content and resource generation. In order to do this, the specification for a linguistic content database which could be used to produce many types of customised resources in varying formats (paper-based or digital) was developed. The specification requirements were that the database content should be curriculum-based, localisable and reusable (through localised settings).

A secondary aim was to compare the features of the raw linguistic content to other types of language learning content, to characterise it and to examine whether it had features of a controlled language.

4.2.1 Phases of work

In order to develop the database specification, a prototype database was designed and developed for the context outlined above through four phases of work. These phases of work included a classroom-based data testing and gathering stage with two primary school classes over two and a half years as described. The four stages of work included:

(I) Preparation, classroom-based research and database specification

(II) Database analysis

(III) Database development and population

(IV) Database proof of concept resource outputs

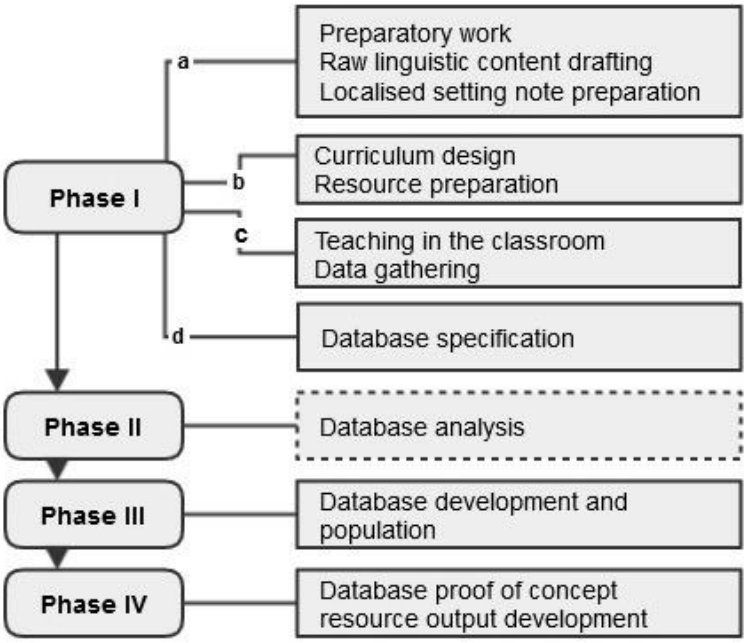


Figure 4.6 Research design overview

Phase II of work – Database analysis – did not directly relate to the development of the database, but was required to examine the secondary aim outlined above. This is indicated by the dashed line of the flowchart in Figure 4.1 outlining the overview of the research design. The database analysis was placed in this position as it examines the raw linguistic content *before* it is added to the database under construction (in its v1.0 format).

The goals of Phases I, III and IV are to progress the specification of the linguistic content database, and the development and population of the prototype database. The goal of Phase II is to characterise the raw linguistic resources and compare them to other types of language learning resources. The processes and methodologies employed in each stage are outlined across the coming sections, Section 4.3 to 4.6. The overall design methodology has its basis in Design-Based Research (DBR).

4.2.2 Methodological choices

This research crosses methodological boundaries and is in essence, a research and development project. Through its division into four phases of work, the appropriate methodologies can be divided across these phases:

- Phase I: Design-based research
- Phase II: Computational Linguistics (CL)/Natural Language Processing (NLP) techniques
- Phases III and IV: Developmental work

The following sections describe the methodological choices which were made for each of the phases of work. Phase I was the only phase through which the children in the primary school classroom were involved.

Design-Based Research

The research reported through Phase I in this thesis has its basis in Design-Based Research (DBR). Design-based research (DBR), sometimes called design research, design experiments or development research, is a mixed-methods approach which involves using both qualitative and quantitative research methods (Anderson & Shattuck, 2012). Using a mixed-methods approach or research design "*provides a better understanding of research problems than either approach alone"* (Creswell & Plano Clark, 2007, p. 5). DBR provided the flexibility to cross qualitative and quantitative boundaries and to include methods from a variety of disciplines (e.g. CL, SLA and MT). DBR has been defined by The Design-Based Research Collective (DBRC) as "an emerging paradigm for the study of learning in context through the systematic design and study of instructional strategies and tools" (DBRC, 2003, p. 5). By its very nature, DBR allowed each of the different types of data required to be recorded in the most appropriate way

Some definitions of DBR (e.g. Anderson & Shattuck, 2012) do not support the teacher also being the researcher as Action Research does. Other definitions of DBR, allow for instances where the researcher "assumes responsibility for instruction" (Cobb et al, 2003, p. 9). The goal of the DBR in this case is gathering data about the localised settings in the local context and an iterative evaluation of the raw linguistic content. The use of DBR as an overarching methodology also encompasses and feeds into one of the outcomes of this research, the untested global development model (Tailor) for the development of a reusable teaching and learning resource database, rather than providing a solution to a local need, as is the case in Action Research. Both DBR and Action Research are suitable for carrying out research in educational settings.

The use of DBR in this research is to iteratively test and evaluate the raw linguistic content developed in Phase I, rather than directly investigating the use of the end product of the research (the linguistic content database with localised settings). In addition, knowledge about and evidence of the localised settings are also gleaned through DBR in the school-based research. Meta-analyses of DBR have revealed that generally, DBR has been used to study theories about the means for supporting learning (Cobb et al, 2003), educational interventions (Anderson & Schattuck, 2012) or educational innovations (DBRC, 2003). Theories focusing on the means for supporting learning can include elements such as "material artefacts, teaching and learning practices and policy levers…" (Cobb et al, 2003, p. 10). Using DBR in this research case focuses on the first of these – material artefacts. Shavelson, Phillips, Towne and Feuer (2003) also support DBR being "utility orientated", among other possibilities like interventionist and theory-

driven," with the intent on improving the effectiveness of instructional tools to support learning" (p. 26).

The use and definition of localised settings to define the context of the research reinforced the view described by Cobb, Confrey, diSessa, Lehrer & Schauble (2003) that DBR can never account for all elements of a complex learning ecology, but must "distinguish in the specification of the design between elements that are the target of investigation and those that may be ancillary, accidental or assumed as background conditions" (p. 10).

CL/NLP Study

The CL/NLP study aims to examine and characterise the language contained within the raw linguistic content. Two other datasets are used as the basis of comparison. Section 3.3 described the analyses which can be carried out from the SLA and CL domains. Controlled language features can be tested for and confirmed or otherwise in a text. The main aim of the CL/NLP study is to confirm whether the raw linguistic content exhibited characteristics of a controlled language and to compare and contrast its makeup to the other two datasets.

Part of speech (POS) tagging is a CL technology which assigns a part of speech to each token in a text. The TreeTagger (Schmid, 1994) part of speech tagger was used in this research. The POS tags were used to isolate certain parts of speech so that features of controlled languages could be identified.

Matrices from SLA can be used to examine complexity, accuracy and fluency (CAF) in a text. Some of these features are relevant to examining controlled language features, and can be used in a cross-disciplinary fashion to do so. Complexity measures are applied to analyse the grammatical and lexical complexity of the raw linguistic content and two additional datasets.

Development work

Phase III requires the design, development and population of an XML database (linguistic content database) which houses the raw linguistic content and their topic-associated localised setting content. Phase IV requires the manipulation of the database and the development of sample proof-of-concept outputs which exhibit how the database can be used as a source of content to develop template-based teaching and learning resources as outputs such as handouts, topic-based lexica and teacher reports for planning. XML, PHP, DTD and HTML were the technologies employed to manipulate the database and develop these outputs.

4.3 Phase I: The Irish primary modern language classroom – preparation, classroom research and database specification

This section outlines the methodological background to Phase I of the work. This phase of work is the only one of the four phases which involved participants. Participant details are described along with the sub-phases including

- I.a: preparatory work before entry into the classroom, raw linguistic content drafting and preparation of localised setting notes on linkage, integration and Language Awareness
- I.b: curriculum design and resource preparation
- I.c: teaching in the classroom and data-gathering
- I.d: database specification.

4.3.1 Participants and duration

The school-based research was conducted in a senior all-boys primary school (second class to sixth class). The segregation of children into male and female schools is a normal occurrence in Ireland, especially in longer established schools. The school is designated as urban disadvantaged, which means that not all children gain full advantage of the education system due to social or economic disadvantage. As such, the school benefits from the Department of Education and Skills (DES) programmes under the *Delivering Equality Of Opportunity In Schools* (DEIS) *Action Plan for Educational Inclusion* (DES, 2005b). These programmes, initiatives and interventions under the Integrated School Support Programme (such as the School Meals Programme, the School Completion Programme and the Reading Recovery Intervention) are designed to try to address elements of educational disadvantage.

Some of the features of educational disadvantage which are targeted in the DEIS Action Plan (DES, 2005a) are outlined in Table 4.8 below alongside the features of educational disadvantage observed in the participating class groups. A focus on disadvantage was not delimited in the localised settings, but it inevitably had an impact on the school-based research. This feature was seen to be too involved for the time available and overall focus of the research. Some features of educational disadvantage influenced the speed at which the German content was covered and trialled, the levels of revision required and the supplementary teaching and learning which was required to teach certain elements of the German content (e.g. revision of what a noun, verb and adjective are). In acting as a baseline for the teaching, learning and trialling of the German raw linguistic content, the local context provided a good and sound test bed. No relevant prior knowledge for learning German (e.g. of parts of speech, knowledge of English and/or Irish) was assumed. This means that all relevant (prior-) knowledge was accounted for in the final trialled

version of the raw linguistic content and can be customised by the teacher in the classroom, depending on the level of ability of the children in his/her class.

Features of Educational Disadvantage (DES, 2005b): ***Difficulties with…***	**Features of Educational Disadvantaged observed and recorded in the participating school**
Literacy and numeracy	Relatively *low levels of literacy* in English and Irish
Attendance	Relatively high levels of *absenteeism* for approximately 20% of class group R and class group S Children persistently arriving *late* to school
Educational progression	Non-completion of *homework*
Retention	Low levels of *retention* Need for constant *revision*
Attainment	Low levels of *attainment* across subjects Need for constant *push* and innovation to retain interest Non-assumption of *prior knowledge*

Table 4.8 Features of educational disadvantage observed in the participating school

One class of 18 children (with an additional child between 17/2/'04 and 30/6/'04, and another additional child between 16/09/'04 and 28/06/'05) (*Class R*) participated in the research over a two and a half year period. A further class of 14 children (*Class S*) participated in the research for a seven month period. Four primary school levels, (1) infant classes, (2) first and second classes, (3) third and fourth classes and (4) fifth and sixth classes are used to delineate the PSC content. The classes involved in the MLPSI, for whom the modern language curriculum documents (NCCA, 1999, 2001) were developed, were the upper level of primary schools, fifth (age 11-12) and sixth class (age 12-13). At the outset, the initial class involved, Class R, was a fourth class (aged 10-11) but the research duration did allow this class group to be followed into their fifth and sixth classes of education (2004/2005 and 2005/2006 respectively). Class S was a fourth class during their time of participation. The younger group in this instance (Class R in 2003/2004 and Class S in 2005/2006) had a negligible impact on the level or types of raw linguistic content being developed, as the duration of time spent in the primary classroom allowed access to these classes as they progressed in their school years.

Duration	Class Group	No. of participants (N)	Class and age	Contact time
23 September 2003 to 17 June 2004	R	18 (+1 from 17/02/04-30/06/04)	4th class Age 10-11	2 classes per week 45 minutes per class 10.00-10.45am
16 September 2004 to 28 June 2005	R	19 (+1 from 16/09/04-28/06/05)	5th class Age 11-12	2 classes per week 45 minutes per class 9.15-10.00am
26 September 2005 to 8 March 2006	S	14	4th class Age 10-11	1 class per week 45 minutes per class 9.15-10.00am
	R	10	6th class Age 12-13	1 class per week 45 minutes per class After school hours – 2.40-3.25pm

Table 4.9 Research duration and participants

The selection of the participants was governed by the teacher of the classes involved. The classroom research was conducted in conjunction with one teacher and her class for any one year over a period of two and a half years (nearly three full school years). Class R was taught by this teacher during the 2003/2004 and 2004/2005 school years. The participating teacher was assigned a new class for the 2005/2006 school year (Class S). Those children from Class R who were interested in continuing on with their German classes and with the research were facilitated in doing so after school hours (hence the reduced number of participants for Class R documented in Table 4.5 below for the 2005/6 school year). The participating teacher remained in class during all of the research German classes (for school-based insurance reasons). Table 4.9 below summarises the research duration and participants at each stage.

4.3.2 Ethical considerations

The classroom-based portion of the research was approved by the Dublin City University Research Ethics Committee retrospectively in 2005. Formal ethical approval was only established as a requirement by the university for research students during this time. It was classified as a low-risk project. The approval provided was to allow for working directly with the children and recording consenting (by parent and child) children's errors from classroom observation, written work and oral examinations (see informed consent form in Appendix B). Parents and children were made aware that data gathered as above would be used to inform this research and also feed back into the materials children were using daily in the classroom. All children and their parents/guardians (across Class R and Class S) consented to each child's participation in the research project.

4.3.3 Preparatory work, raw linguistic content drafting and compilation of relevant areas of linkage, integration and language awareness

Preparatory work and raw linguistic content drafting were carried out before commencing the research in the classroom. Adequate preparation ensured that the time spent in the classroom could be focused exclusively on data gathering and teaching.

(1) Preparatory work

Once the context for the research had been established, the contextual requirements and parameters were defined according to the localised settings (as outlined in Section 4.1.3 above). Background reading and research into each of the identified localised settings were carried out (as summarised in Section 4.1.4 above). At the same time, a primary school was sought within which the classroom-based research could be carried out. A local teacher agreed to participate in the research and assign two 45-minute sessions per week of her teaching time to German language teaching.

The problem space for this primary school classroom and its children was outlined above in Section 4.1.3. Literature reviews of each of the localised settings allowed the smooth transition between and coexistence of roles as researcher and teacher in the classroom. They also allowed the drafting of notes on linkage, integration and language awareness by topic which sit alongside the topics of the raw linguistic resources (see below). Finally, these literature reviews allowed data gathering instruments to be devised for use in the classroom. The areas of investigation and the instruments used are outlined in Section 4.3.5 below.

(2) Content drafting

The development of the raw linguistic content followed a series of steps (inputs and processes). The outputs trace the development of the raw linguistic content in its various stages of development. These steps are summarised in Figure 4.7 below and will be considered and discussed across the coming sections. The dashed lines of the figure indicate where human variance is at play, with the full-lined processes being fully transferrable to other contexts.

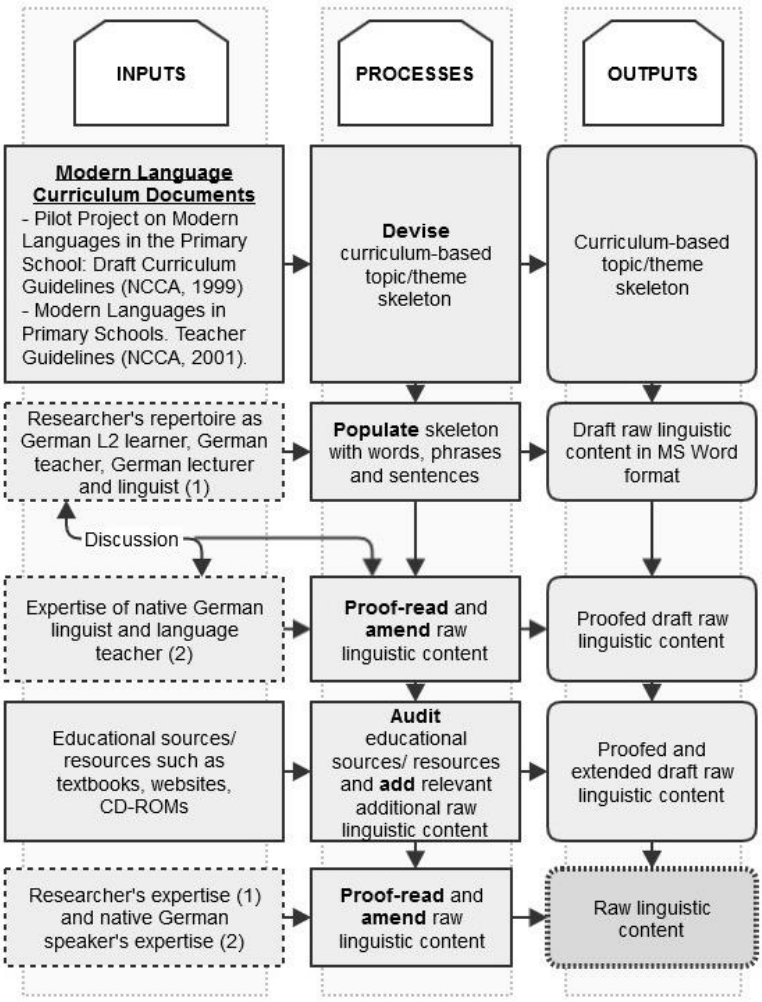


Figure 4.7 Phase I: Content drafting

Examination of the modern language curriculum documents and extraction of the core topics, themes and learning goals

The first step was a consideration and consultation of the modern language curriculum documents which comprise *Pilot Project on Modern Languages in the Primary School: Draft Curriculum Guidelines* (NCCA, 1999) and *Modern Languages in Primary Schools. Teacher Guidelines* (NCCA, 2001). The core topics, themes and learning goals were extracted from these documents to form a skeleton structure within which the raw linguistic content could be added. While the language topics and language functions in themselves would have represented a suitable format to follow for the skeleton structure, these topics were arranged into the order presented in Table 4.10 below, as it was a better method to chronologically and gradually present the information in the classroom (starting with more basic content and progressing to more difficult content). The language topics would be presented in order or in a non-linear fashion, with fillers (e.g. days of the week) and classroom language (e.g. "May I go to the toilet please", "I don't understand", "can you say that again please?") being used at any appropriate stage. It also allowed topics to be revisited more often as a requirement for the local research school

being a DEIS school (e.g. (1) General introductions, (4) General introductions – More Advanced I).

Language Topics	1. General Introductions	
	2. Parts of the Body	
	3. Numbers 1-20	
	4. General Introductions – More Advanced I	
	5. General Introductions – More Advanced II	
	6. Finding Out What Things Are Called	
	7. Family I	
	8. Weather	
	9. Family II and Describing Yourself I	
	10. Describing Yourself II	
	11. School	
	12. Pets	
	13. Pastimes and Hobbies	
	14. Numbers 20+	
	15. Time	
	16. Birthdays and Date	
	17. Food	
	18. Fillers	Days of the Week, Months of the Year, Shapes, Colours and Seasons
	19. Classroom Language	

Table 4.10 Skeleton structure of the raw linguistic content

Extension of the skeleton curriculum topics and themes to words, phrases and sentences

Once the skeleton structure of the raw linguistic content had been finalised, the next stage of work involved working all of these topics through from beginning to end and populating them with content. This initial work was carried out from the researcher's own repertoire as German L2 learner, German teacher, German lecturer and linguist. The language register, language progression over time and appropriateness of content to the learners would be taken into account through the classroom research.

The initial version of raw linguistic content was drafted using a table-based format in MS Word to allow for transferability between the researcher, the participating teacher and a native German speaker (a linguist and language teacher) who would proof-read the initial draft of the raw linguistic content. The section headings were all written in English and the core language in German only (without any English translations). Table 4.11 below provides a sample of this content.

Section One: General Introductions Language:	
1. Greetings Hallo! Guten Morgen Guten Tag Guten Abend Gute Nacht Schlaf gut	 Auf Wiedersehen Tschüs Bis bald Bis morgen/Donnerstag…
2. How are you? Wie geht's? Gut, danke! Sehr gut, danke! Nicht so gut Schlecht Mir geht's gut/sehr gut…	3. What is your name? Wie heißt du? Ich heiße _____

Table 4.11 Excerpt from the raw linguistic content after the initial addition of words, phrases and sentences

Once the content was drafted, it was proof-read by the researcher to ensure (1) it met the requirements of the curriculum documents and (2) that the language was grammatical and free from errors. The content was then also proof-read by the native speaker. The proof-reader suggested a few changes and additions which included alternative ways to phrase some sentences, alterations to the phrasings to use the more common form of a phrase and additions of vocabulary which was specific to the research context (e.g. the correct translations for two of the PSC subjects). Following discussions with the proof-reader, some of the alternative phrasings suggested were deemed to be too formal for children speaking to one another or indeed to adults. In these instances, the original format was left. Where a suggested format was suitable, it was included alongside the original form.

Audit of relevant texts and electronic language learning resources for the local context and addition of content where appropriate

Once the initial draft of the raw linguistic content was written and proofed, additional raw linguistic content from relevant sources was gathered and added where appropriate. These relevant paper- and e-sources included German language learning textbooks for children, secondary school children (beginners) and adults (beginners), online sources such as German language learning sites (particularly geared towards children or young adults) and German language learning CD-ROMs which are targeted at children. A listing of the resources consulted is outlined in Appendix C, page 287.

Relevant content was searched for across the language learning resources by topic (following the topics of the raw linguistic resources being drafted). When the textbooks were consulted for example, each topic considered had relevant content spread across a large number of textbook pages. The language to be taught in these textbooks needed to be placed in context and within the scope of an activity or learning task (or follow the Present, Practice and Produce Method (PPP), or the alternative Observe, Hypothesise and Experiment), so would have taken up more page space than the same content in the raw linguistic content. The extent of this spreading was greater than expected. In contrast, the content outlined around each topic in the raw linguistic resources is condensed.

This phase of work involved the addition of many nouns (as additional items relevant under the identified themes and topics), and alternative phrases (listed in Table 4.12 below). The total additional word count for this phase reached 205 words (resulting in a total word count of 2,216).

These additions were made for a number of reasons. Some of these additional words provide a good opportunity to explore cognates, which are words from different languages which look and sound the same (e.g. Die Melone [*melon*], Die Aprikose [*apricot*], Jazz, Talk-Shows). Cognates are relevant to the chosen localised setting of Language Awareness. Other words were added so that children could link and integrate with other areas of the PSC in being aware of writing genres (romantische Bücher [*romance novels*], Biographien [*biographies*], Autobiographien [*autobiographies*], Reisebücher [*travel books*], Zeitungen [*newspapers*]) and different types of music (e.g. klassische Musik [*classical music*], Instrumentalmusik [*instrumental music*], Blues-Musik [*blues music*], Jazz), for example. Other additions were included as they provided an alternative way to phrase something. Others were oversights, but which are very relevant to the Irish content (e.g. Die Sommersprosse(n) [*freckle(s)*]).

The additions only made up ca. 10% of the new whole of the raw linguistic content. However, the result of consulting other resources as part of the development process *cautions against* the use of introspection and using your own 'internal' repertoire as a sole means of authoring content as the omission of relevant words, phrases and sentences is unavoidable.

Topic	Additions	
Section One: General Introductions	Bis morgen/Donnerstag…	
Section Two: Parts of the Body	Der Daumen Die Zehe	
Section Six: Finding out what things are called	Ein/ Der (Wasser)Hahn Eine/ Die Spüle	Ein/ Das Spülbecken
Section Seven: Family	Der Mann Der Freund Der Junge	Die Frau Die Freundin Das Mädchen
Section Eight: Weather	Wieviel Grad haben wir heute? Der Niederschlag	
Section Nine: Family 2 – Describing yourself and your family	Komisch Stark	
Section Ten: Describing yourself 2	Die Sommersprosse(n)	
Section Eleven: School (including uniform/clothes)	Durchfallen Die Karte Der Stundenplan Der Direktor Die Garderobe Die (Unterrichts)stunde Die Direktorin Einen Blazer	Shorts (pl.) Turnschuhe (pl.) Der Gürtel Der Knopf Der Schnurrsenkel Der Stiefel (sing.) Der Trainingsanzug Die Sonnenbrille (sing.)
Section Twelve: Pets	Ja, ich habe ein Pferd einen Frosch einen Igel eine Schlange	eine Schildkröte eine Schnecke eine Spinne ein Pferd
Section Thirteen: Pastimes and hobbies	Was machst du abends? Ich gehe gern kegeln Ich bastele gern Ich spiele Volleyball Ich spiele bowling Ich sammle/tausche Briefmarken Ich reite gern Ich spiele lieber mit dem Computer Ich spiele Geige klassische Musik Instrumentalmusik	Blues-Musik Jazz Manchmal gehe ich zu einem Konzert Ich finde sie prima romantische Bücher Biographien Autobiographien Reisebücher Zeitungen Talk-Shows
Section Fifteen: Time	Wie spät ist es bitte? Am Wochenende stehe ich um 10 Uhr auf.	
Section Seventeen: Food and shopping	Der Spinat Der Rosenkohl Die Melone(n) Die Kirsche(n) Die Aprikose(n)	Die Pflaume(n) Der Pfirsich(e) Das Obst Das Gemüse Das Fleisch
Classroom Language	Darf ich bitte auf die Toilette gehen? Könnten Sie bitte mal herkommen? Ich kapiere das nicht. Wer ist das? Würfle!	Mein Hund hat die Hausaufgaben gefressen! Vielleicht Später Jetzt

Table 4.12 Words, phrases and sentences added to the raw linguistic content following consultation of other resources

The final raw linguistic content was again proof read by the same native German speaker. No further amendments or additions were suggested. The full listing of the raw linguistic content is provided in Appendix D, page 288, with additions from alternative resources highlighted in yellow.

(3) Consideration of notes on localised setting content: linkage, integration and language awareness

A review of **integration and linkage** identified cross-over and potential areas for integrating strands and strand units of the PSC within the German language lessons, as well as areas to link within the German curriculum itself. The focus was on relevant strands and strand units of the PSC, rather than going to the deeper level of objective. Much of the relevant areas of integration and linkage related not only to content (raw linguistic content) but were also relevant to the teaching activity which would be carried out. Any areas identified which were relevant to an activity (e.g. drawing your house was relevant to Visual Arts and listening to and learning relevant songs was relevant to Music) rather than the raw linguistic content were omitted. As a result, PSC subjects which are more skills-based (e.g. Mathematics, History, Geography and Science) or concepts-based (e.g. Visual Arts, Music) tended to have few listings within the integration topics. This resulted in the most relevant areas of integration stemming from the other language subjects – English and Gaeilge/Irish, or subjects which had a direct correlation to the topics within the raw linguistic content (e.g. SPHE, History, Geography).

Areas for linkage were backwards and forward compatible, that is, they took account of the relevance of each section to future and past sections. In practice, this meant that links were made with content that had been covered in previous sections, but also with content which would be covered in later sections. This linkage information formed a quasi-continuum of learning, which recorded areas which were related to one another from previous and future sections. This accounted for the non-linear fashion through which the topics could be followed in the classroom. An excerpt of the areas of integration and linkage are provided in Table 4.13 below. The full listing is provided in Appendix E.

Section	Linkage and Integration Notes
1. Greetings	Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >téamaí: Mé féin *Irish>Listening, speaking, reading, writing>themes: Myself* • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • History>Myself and my family/local studies>Myself • SPHE>Myself>Self-identity • SPHE>Myself and Others>Relating to others Linkage • Section Four: General Introductions – more advanced 1 • Section Five: General Introductions – more advanced 2 • Section Six: Finding out what things are called • Section Seven: Family • Section Twelve: Pets
2. Parts of the body	Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >téamaí: Mé féin */ Irish>Listening, speaking, reading and writing>themes: Myself* • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • SPHE>Taking care of my body>Knowing about my body • Science>Living things>Myself/Human life Linkage • Section Twelve: Pets
3. Numbers	Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht > */ Irish>Listening, speaking, reading and writing>* • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • Mathematics>Number>Counting and numeration and comparing and ordering • Mathematics>Measures>Money Linkage • Section Four: General Introductions – more advanced 1 • Section Five: General Introductions – more advanced 2 • Section Seven: Family • Section Fourteen: Numbers 20+ • Section Fifteen: Time • Section Sixteen: Birthdays and date • Section Seventeen: Food (and shopping)

Table 4.13 Excerpt from integration and linkage notes for Sections 1, 2 and 3 of the raw linguistic content (Note: The use of the '>' indicates how the PSC content goes from Subject>Strand>Strand unit)

A review of **Language Awareness** and the relevant areas of the curriculum (English, Gaeilge and the modern language) revealed relevant areas of focus, similarity and difference across the languages. The raw linguistic content was translated into English and Irish to also assist in

identifying some relevant areas of Language Awareness (e.g. similarities and differences in words, sounds and structures). These were documented in the raw linguistic content (Raw linguistic content version 1.1 – see Appendix E for the full listing) and an excerpt is provided below in Table 4.14.

Section	Language Awareness Notes
1. Greetings	• Examine word order for naming yourself in Irish differs from German and English – [] is ainm dom, ich heiße [], my name is []. • Examine 'hello' expressions in German, English and Irish. Note 'dia duit' and its response(s). • Examine dialects across the three languages, e.g. in Irish – three ways of asking 'how are you?' - Chaoi a bhfuil tú? Conas atá tú? Cad é mar atá tú? • Examine contractions in English (e.g. how're/how are), Irish (e.g. Táim/Tá mé) and German (e.g. Wie geht's/wie geht es dir) • Examine how questions are asked in each of the languages.
2. Parts of the body	• Examine words which look/sound like the English equivalent (cognates) (e.g. Finger, Fuß, Nase, Haare, Schulter, Hand, Knie). • Examine language families as the reason why there are many English/German similarities and few Irish/German ones • Examine English words which contain German ones/have their root in German – Rucksack/back pack from the German 'Rücken' • Describe nouns – what is a noun? • Examine how all nouns in German receive a capital letter. Examine the use of capital letters in English and Irish. • Examine the use of gender with nouns in German (masculine, feminine and neuter) and Irish (masculine and feminine)
3. Numbers	• Examine words which look/sound like the English equivalent and mean the same thing (cognates) (e.g. Telefonnummer) or Irish equivalent (e.g. acht/a hocht, Telefon/fón/theileafóin). • Examine words which look/sound like an English word, but which do not mean the same/similar things (false friends) (e.g. drei/dry, vier/fear, elf/elf, zwölf/wolf) • Examine language families as the reason why there are many English/German similarities and few Irish/German ones • Examine how numbering schemes work in English, German and Irish – e.g. addition of 'teen', 'zehn' and 'déag' • Examine word order, Irish is a verb initial language (VSO: Verb Subject Object) and English is an SVO language. German is also an SVO language, with SOV subordinate clauses.

Table 4.14 Excerpt from Language Awareness notes for Sections 1, 2 and 3 of the raw linguistic content

Areas of focus for Language Awareness notes usually centred on similarities and differences between the three languages which were lexical (e.g. vocabulary, cognates, false friends) and/or grammatical (e.g. word order, negation, compound words, forming the plural). The former lexical items helped to focus on comparing and contrasting the three languages while the latter also focused on comparing and contrasting but also on areas which require focus when learning different languages. This can be referred to as "learning to learn" (KEC, 2007, p. 32), and also

includes strategies to help learn language(s). Grammatical focus also provides students with an opportunity to 'problem solve' and figure out how the grammatical system works across languages in their environment (e.g. negation, plurals).

The potential areas of integration, linkage and language awareness were topic-based and sat alongside the raw linguistic content. They could be considered in each case as the content for teaching was selected each week.

4.3.4 Curriculum design and resource preparation

Curriculum design

The teaching and learning which took place in the classroom was bounded by the ethos of the modern language curriculum documents (NCCA, 1999, 2001). As described, these national curriculum documents follow a communicative language learning approach. Language functions such as "giving and seeking information" (NCCA, 1999, p. 14) also allowed more task-based learning to be integrated.

The contents of the draft raw linguistic content are based on the curriculum topics. Their content progresses from first lesson introductions to more advanced topics which build on what has been learned previously. The content could therefore be followed in a chronological order or accessed in a non-linear way depending on the requirements of the children in the classroom. The localised setting information sits alongside the raw linguistic content, and is also topic-based to correlate to the raw linguistic content. Selections of raw linguistic content topics and language, also include the relevant localised setting information.

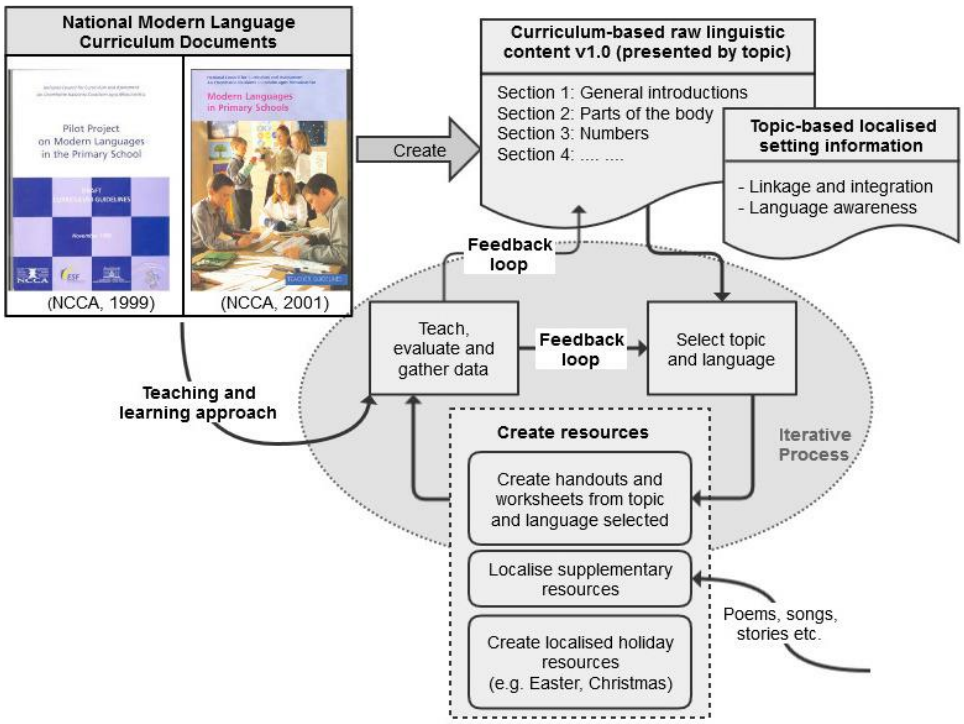


Figure 4.8 Curriculum Design

The curriculum design comprises three iterative processes – (1) Content/topic selection, (2) resource generation and (3) classroom teaching, evaluation and data gathering. The iteration occurs through evaluation and feedback being gathered through teaching in the classroom, which feeds back into the content selection and resource generation for subsequent lessons. A further feedback loop provides evaluative data which feeds back into the development of the raw linguistic content. This feedback will be discussed in Chapter 5. The curriculum design processes are already ring-fenced by the overarching modern language curriculum documents which broadly specify what is to be taught and the teaching approach to be adopted. The processes and surrounding context are indicated in Figure 4.8.

Resource preparation

Materials or resources (handouts and worksheets) were manually prepared to present the raw linguistic content on a weekly basis and used the relevant topic title or subtitle (e.g. greetings, months of the year) for ease of reference. These were developed on a topic by topic basis, where relevant raw linguistic content and associated localised setting information for linkage, integration and language awareness were selected in unison. Supplementary materials (e.g. poems or songs) were also sourced to reinforce the language which had been taught and encountered. In some instances, additional materials were developed where there was a need to revise and reinforce a topic more thoroughly or to recognise a holiday season (e.g. Christmas, Easter).

The handouts and worksheets (based on the raw linguistic content Version 1.0 with LS notes) were developed in black and white (or where colour was used, it could be printed/copied in greyscale) to allow for photocopying, contained images and strived for a good layout and presentation. Equally, it was important to try to maximise the use of the space on the page to reduce photocopying costs (an important factor in the participating school). However, the quality was far removed from a professionally produced colour textbook.

The handouts were used to present the teaching content during teaching and learning in the classroom. They served as a reference point for the children during task-based activities and as a record of the topics being covered in class. They were also used to assist with their practice at home. The corresponding worksheet activity was used as a reinforcement and practice activity for school or home. Sample handouts and worksheets illustrating the different types of resources which were created (based on the raw linguistic content, localised supplementary resources and holiday-related resources) are provided in Appendix F. The next stage which employed the handouts and worksheets was teaching in the classroom.

4.3.5 Teaching in the classroom and data gathering

The data gathering element of this sub-phase drew on Hubbard's (1996) evaluation model to account for learner fit of the raw linguistic resources and localised setting notes which had been developed. Hubbard's (1996) Evaluation Module is the inverse of his Development Module. Learner and teacher fit are determined by checking in that all of the design criteria were met in practice.

(1) Teaching in the classroom

Having taken on the role of visiting German teacher and researcher in the classroom, the research was facilitated through the course of German teaching, learning and assessment. As outlined above, literature reviews on the selected localised settings contributed to the preparatory work which was carried out before entry into the classroom. Examining the Primary School Curriculum and the modern language curriculum documents were equally important in preparing to operate as a teacher in the classroom as well as inform the initial development of the raw linguistic content and associated linkage and integration and language awareness notes.

Once the relevant handout and worksheet had been created (manually), its content was presented to the children in the classroom. Children would then be provided with an opportunity to practise using the new or revised language through communicative and task-based activities like role-plays, dialogues and questions and answers. Games, songs and poems were frequently used activities and resources, which included bingo for practising number recognition, Simon says

(Pumpernickel sagt) for listening and comprehension of parts of the body (total physical response) and team-based races to pin the correct body part label on the body. With a focus on integration, teaching activities were also aimed at complementing those carried out for other subjects of the PSC with the class teacher. As the children's German repertoire built up, lessons commenced with a focus on class news and the weather (as done in Irish also), and then progressed on to new and revised material presentation and practice activities. Follow-up activities were usually assigned towards the end of the lesson for completion at home.

As described above in Section 4.3.1, the children participating in the research were attending a school which was designated as a DEIS school. Some children would arrive late to class, some may not have completed homework assignments, some had low levels of retention and some had very low levels of literacy. Each lesson focused on a thorough revision of language learned which was relevant to the lesson underway. No previous knowledge was assumed and all grammar-based focuses were explained from a zero base level (e.g. when encountering articles and nouns, the children were also taught what these parts of speech were and how they were used before progressing on to their relevance to the lesson). While a certain amount of reading and writing was required, the emphasis on communicative and task-based activities assisted the learning and reinforcement of German language without placing an overemphasis on the written word.

(2) Gathering relevant data

The focus of the data gathering was on the raw linguistic content in the handouts and worksheets and the associated localised settings (integration, linkage and language awareness) which were used and presented during the teaching activities. Learner error analysis was also a focus of data gathering. Data gathering was confined to an examination of these areas and not to the teaching activity or other elements which created part of the lesson-whole or school-based context.

The questions associated with each of the areas of focus are outlined below in Table 4.16. Each of the following subsections will describe the processes in more detail. Table 4.15 indicates the data gathering timeline and areas of focus during the 2.5 years in the classroom.

Data sought	Timeline (by school year)
Evidence of any changes required to German raw linguistic content [pace, content, presentation]	2003-2005
Evidence of appropriate links and areas of integration between the PSC and the German raw linguistic content	2003-2005
Evidence of types and frequency of learner errors	2003-2005
Evidence of appropriate elements of Language Awareness for Irish, English and German	2003-2005
Evidence from trialling the complete German raw linguistic content, evidence from testing the completeness of data gathered on learner errors, Language Awareness and PSC links and integration (consolidation)	2005-2006

Table 4.15 Data sought and associated timeline

Raw linguistic content

The first and second years of data gathering in the classroom (2003-2004 and 2004-2005 school years) allowed the developed raw linguistic content to be trialled and tested in a classroom setting with Class Group R. The third year of testing allowed the content to be tested with a new class group, Class Group S

Existing classroom observation schemes can be sourced and used (e.g. Spada and Fröhlich's Communicative Orientation of Language Teaching (COLT) Observation Scheme, 1995). "They differ in several respects, including the number of categories they contain, whether they focus on qualitative or quantitative descriptions, and whether they are used throughout a lesson or on selected samples of classroom interaction" (Lightbown & Spada, 2006, p. 114). However, existing observation schemes, while providing a starting point and basis for the development of research-specific observation schemes, will generally be too restrictive, or somewhat irrelevant, in that they have been developed for alternative purposes (e.g. to observe pre-service language teachers during their teaching practice). Similarly, existing observation schemes assume that the observer is not participating in the classroom interactions, as teacher or researcher.

Focus of data gathering	**Areas of investigation**	**Data-gathering Instruments used**
Raw linguistic content	**Content** • Is there any relevant language (words, phrases, sentences) which is missing from the handout/worksheet • Is there any superfluous or irrelevant language on the handouts/worksheets? • How relevant is the content to the children? (not relevant / relevant / very relevant) • Were any difficulties encountered with the language presented (e.g. vocabulary, pronunciation)? If yes, provide details. **Resources/presentation** (of relevance to the outputs of the linguistic content database) • How clear is the handout/worksheet (not clear / clear / very clear • How appropriate is the handout / worksheet (not appropriate / appropriate / very appropriate)? • Any changes required? If yes, give details • Researcher's ancillary notes	Classroom observations noted on structured feedback form and relevant handout and worksheet
Integration and linkage	• Were the integration and/or linkage topics(s) identified appropriate? • Did any other topics for integration or linkage arise?	Classroom observations noted on handout and worksheet Reflective cues used indirectly after the lesson
Language awareness	• Were the Language Awareness topics(s) identified appropriate? • Did any additional LA elements emerge through the course of the lesson?	Classroom observations noted on handout and worksheet Reflective cues used indirectly after the lesson
Learner error analysis	Type and frequency of learner errors. (1) Learner errors identified (2) Learner errors described by classifying them according to types	Classroom observations noted on closed-ended checklist of learner error domains and categories. Hard copy written and recorded oral tests were analysed according to the same checklist.

Table 4.16 Summary of data sought and instruments employed

During each lesson, the raw linguistic content was examined for its appropriateness in content and means of presentation. The former was directly relevant to the raw linguistic content and the latter to the presentation of the proof of concept outputs of the linguistic content database. I used

a structured feedback form for noting observations relating to the raw linguistic content and the resources/presentation of the raw linguistic content (see Appendix G). The structured form used the relevant handout and worksheet as cross referent for each topic within the raw linguistic content. The questions of focus are outlined in Table 4.16 and in Appendix G. The content focused questions examined the comprehensiveness of the content, its suitability for the children and any difficulties encountered in its use. The presentation focused questions studied the clarity and appropriateness of the handouts and worksheets and whether any alterations were required. The final field on the structured feedback form allowed any ancillary notes to be recorded.

Integration and linkage

Any observations relating to integration or linkage were noted on the relevant handout or worksheet for any one class, during the class. The handout and worksheet were used as they contained the relevant information to which information could be added, and/or subtracted. The instruments used were refined after an initial trial period and were used to focus reflection on the relevant handout and worksheet but did not require direct completion on a separate form. This ensured that data gathering did not overtake the teaching and learning which was also taking place. The cues were trimmed to only include those which were most relevant to the raw linguistic content and localised settings, as follows:

- Were the integration and/or linkage topic(s) identified appropriate?
- Did any other topics for integrating or linking arise during the class?

These cues were typed up alongside the language awareness cues (see below), laminated, and attached to the inside of the lesson/research notebook for consideration after each lesson. Any direct changes necessary were made to the handout/worksheet and subsequently the raw linguistic content.

Many potential areas for linkage and integration were identified, more than could be tested within the time limits of the classroom research. As such, the more relevant areas of integration were selected according to their relevance to the *teaching* going on in the classroom and *school* at any one time (and which also reflected those outlined in advance of entering the classroom). The former were established through evidence of the print-rich environment in the classroom and through discussions with the classroom teacher. The latter were established by the time of year and holidays which were taking place. While not directly related to the raw linguistic resources, their inclusion in the classroom allowed for a more balanced and authentic learning experience for the children involved.

Relevant areas of linkage which were examined were applicable through the links made in the teaching topic for any given lesson. That is, when new material was being presented, any

relevant prior knowledge from previously covered topics was also drawn upon for revision and reinforcement.

Language Awareness

Following the same pilot phase working with reflective cues to gather data on Language Awareness, the data gathering instruments were trimmed and simplified to ensure that the data gathered was relevant to the raw linguistic content and did not dominate the German lessons. Similarities, differences and relevant areas of focus, grammatical or otherwise (e.g. grammatical differences, like the adjective/noun positions for German, English and Irish) were recorded.
The reflective cues were trimmed to refine them to:

- Were the LA topic(s) identified appropriate?
- Did any (additional) LA elements emerge through the course of the lesson?

These reflective cues were included alongside those for linkage and integration on the laminated set of questions in the lesson/research notebook. Areas of Language Awareness were addressed through the course of teaching, but similarly to the examination of linkage and integration, there was insufficient time in the classroom to address all of those identified.

Learner error analysis

The initial work in examining learner errors required a closed-ended checklist of relevant learner error categories to be compiled. Compiling an error checklist which would be relevant to the data emerging from the classroom required analysis of existing error taxonomies, as well as an examination of the types of errors which could emerge. The former were examined in Chapter 3 in respect of the localised setting on error analysis. The latter was drafted from analysing a series of copybooks which children similar to the participants had used in the same setting for noting their German learning and completing German homework and exercises. These copybooks were made available by the participating teacher who had partaken in another German teaching and research project with the researcher during a previous school year.

As the raw linguistic content was restricted to the present tense, active voice and represented a subset of lexis and language (defined by the topics and the raw linguistic content within each topic), some typical error domains and categories such as Granger's (2003) Register>Syntax (<RSY>) or Grammar>Tense (<G><TPS>) were not relevant to the research classroom. This allowed a reduced set of domains and categories to be considered for the closed-ended error checklist. Learner errors were to be documented from evidence of participating children's written and spoken production and through classroom observation and post-production analysis. As such, the checklist needed to be suited to written (e.g. to include typography as in Granger, 2003) and spoken production (e.g. to include pronunciation errors as in Mackey, Gass &

McDonagh, 2000) as well as being quick and easy to use/record while also teaching. The finalised checklist (see Table 4.17 below and Appendix H for a more in-depth description) was based on the taxonomy discussed in Granger (2003) and Mackey, Gass and McDonagh (2000), with error tags being localised for easier and more intuitive identification. The latter taxonomy focused on the analysis of transcribed spoken language and was relevant to the oral test and work which would be analysed.

Error domain		**Error category**		**Error type**	
<W>	Written form	<FM>	Form	<CA>	Incorrect case (upper/lower)
				<DIA>	Diacritic (missing or erroneous)
				<PS>	Phonetic spelling
				<SE>	Spelling Error (Misc.)
		<PN>	Punctuation	<MP>	Missing punctuation
				<EP>	Extra punctuation
<L>	Language/Lexis	<CS>	Code switching	<ENG>	Use of English word
				<IR>	Use of Irish word
		<CE>	Choice	<IC>	Inappropriate lexical choice
<M>	Morphology	<IN>	Inflection	<PL>	Incorrect plural form/Missing plural marker
				<II>	Incorrect inflection
				<MI>	Missing inflection
<G>	Grammar	<AT>	Agreement	<ArtN>	Article Noun agreement
				<AdjN>	Adjective Noun agreement
				<SubjV>	Subject Verb agreement
				<AdvN>	Adverb Noun agreement
				<PrepArtN>	Preposition-Article Noun agreement
<S>	Syntax			<WO>	Word order
				<MW>	Missing word
				<EW>	Extra word
<U>	Understanding			<Und>	Lack of understanding
<P>	Phonology			<Pron>	Difficulties with pronunciation

Table 4.17 Error taxonomy

If we re-examine Ellis' (1997) four stages of learner error analysis (1) identify the error, (2) describe and classify the error, (3) explain the origin of the errors and (4) evaluate the learner errors, the first two were relevant to data gathering and were noted for addition to the linguistic content database. The third and fourth were relevant in operating as a teacher in the classroom.

Students' **oral errors** noted during their class interaction and responses (through pair work, group work and whole-class work) were noted and recorded by ticking the relevant category on

the closed-ended checklist of learner error domains and categories. Additional ticks were added where an error occurred more than once per student and per session.

An oral test was conducted (also within consent) for Class R in 2004. The oral test covered topics which had been taught in class and took a question/answer format. Students were prepared for the oral test with a revision sheet of questions they would be asked and were asked to prepare their answers to these questions (see Appendix J *Oral Test Information*'). The oral test was recorded onto a dictaphone and transcribed so that errors could be identified.

Students' **written errors** were encountered through their written work in class, German copybook, worksheets and any written tests which were completed. Errors were analysed (document analysis) according to the same closed-ended checklist of errors and were analysed by type and frequency. Written errors were always analysed outside of the classroom after they had been gathered. Once collected from any one class, errors would be noted by type and frequency, and written work would be returned for the next teaching session. In some instances, a copy of the children's written work was retained as evidence and example of the data gathering process.

Once data had been gathered and analysed (see Chapter 5 for findings), the next stage of work was to specify the linguistic content database which would house the raw linguistic resources and associated data relating to each of the localised settings. The data gathered in the classroom impacted the database structure as it guided the level of importance which was attached to the content which was presented and gathered. The next section outlines the resulting database specification.

4.3.6 Database specification

Colpaert (2004) noted that "a great contribution to CALL would be a generic database structure that would guarantee reusability of content" (p. 63). The specification of the linguistic content database had three guiding factors which impacted its structure and needed to be addressed. It needed

(1) to be accessible to the end user for the quick and easy development of resources
(2) to account for the localisation of resources through its structure and content
(3) to allow for reusability and transferability of content
(4) to ensure the separation of content from its product.

The localisation of content was achieved through the development of the raw linguistic resources around the relevant context curriculum, and the inclusion of localised settings. Testing and

extending these elements from research in the classroom added further to the localisation of the content.

The second guiding factor and identified resource challenge – reuse of content can be considered in two ways. The first, is where the localised content is re-employable in alternative settings. Localised settings create the facility for this reuse of content. As specified in Section 4.1.3, localised settings can be switched on or off, depending on the context in which they are being used. The content can be reused in alternative contexts, where some of the localised setting details intersect. The second reuse of content occurs through the strict separation of data from its product, which overlaps with the third factor identified.

The sheer fact that a linguistic content database was the chosen solution for this research, addresses the third guiding factor. Databases by their nature hold content which is accessed or queried via an interface. In this way, the content is retained separately to its product.

The database also needed to account for the information which would be held within it. An analysis of relevant data types and categories was carried out, which listed the content to be stored in the database. Table 4.18 outlines these datatypes.

Data		Data Type(s)
Raw linguistic content		1. Headings (English) 2. Words (German) 3. Vocabulary items (Article Noun pairs) (German) 4. Full Sentences (questions, statements) (German) 5. POS data for all German content (2-4 as listed; tagset) 6. English and Irish translations of the German content (2-4 as listed; without POS details)
Localised Setting (LS) (2):	Linkage	References to internal section links (by heading name) (English)
	Integration	References to Subject>Strand>Strand unit of the PSC (English/Irish)
LS (4): Language Awareness		▪ Sentences referring to raw linguistic content sections (English/Irish) ▪ Sentences referring to relevant themes which are common to the raw linguistic content (e.g. dialects) (English/Irish) ▪ Linguistic/grammar items (e.g. compound words) (English/Irish) ▪ Language learning items (English)
LS (5): Error Analysis		▪ Listing of errors made by type and frequency for each raw linguistic content section ▪ Top 5 errors made in any one section as focus of potential errors for language learners ▪ Integrated language awareness and learner error focus notes, which help to identify pressure points in learning which can be addressed through teaching.

Table 4.18 Data types of relevance to the linguistic content database

The relationships between the different data types also needed to be considered. All data was linked to each other by the section header. For example, Section One on Greetings contained data on each one of the listings in Table 4.1. There were different types of 'sentences' in the German content (words, vocabulary items, full sentences) for which there were English and Irish translations which needed to be connected to (or aligned to) the relevant German content. Creating proper links between the POS-tagged German content and the non-tagged English and Irish translations would be important (e.g. to allow the extraction of lexica).

The linkage references referred to internal sections of the content (as opposed to alternative external content). Integration references referred to external PSC subject references, with the curriculum subject as the guiding data. Language awareness and error analysis content all needed to link to the relevant section heading content and potentially to other sections (where similar content could be encountered).

A series of steps was followed to ascertain the correct database solution:

- Research and select the relevant database format
- Match the required data and structures to the chosen database options available
- Assess conformance to the three resource requirements

- Specify the database content.

As a result of working in the classroom and examining the datatypes which were important for inclusion, the database specification could be finalised. Chapter 7 outlines how the steps outlined above progressed.

4.4 Phase II: Database analysis

This analysis of the raw linguistic content was carried out at the Phase II juncture as it is at this stage that the raw linguistic content is still available in raw text format and has not begun its migration into a database format. The localised setting notes have not yet been added to the content (rather it sits alongside it). As mentioned in the research design overview, the dashed line surrounding this phase of work indicates that the raw linguistic content and/or linguistic content database do not alter during this phase, rather the former is examined and characterised.

The aim of the database analysis was to compare and contrast the raw linguistic content to two other data sources which could be used as language learning content also. A sub-aim was to investigate whether the raw linguistic content represented a controlled language and/or limited domain in its bounding by a national curriculum and restriction to beginner's learner language used by children.

In order to analyse the content of the three datasets, a certain amount of pre-processing was required which made the content more amenable for accessing. POS tags were added to the content, so that certain parts of speech could be isolated and counted. Similarly, the structure of the content was altered to allow for easier access to component parts by the programming scripts used (see Appendices N, O and P).

Custom Perl scripts were written to process the content (Appendix N) and the Treetagger POS tagger (Schmid, 1994) was used to add the respective parts of speech. Ratios from SLA and controlled language features were investigated (as described in Section 3.3). The relevant ratios and features were selected from a lengthy list of possibilities. In some cases, multiple ratios to calculate the same feature were used to allow for balance among the data and the results. Custom Perl scripts were also used to locate and track the relevant content in the three datasets for the feature-based and ratio-based analysis (Appendix P).

Once all of the relevant ratios and features had been tested for, the data was presented in MS Word tables. Chapter 6 describes the analysis of the data and the result of the testing. Once the specification was complete, the next stage was the design and development of the database.

4.5 Phase III: Database development and population

Phases III and IV are described in Chapter 7 together and although they represent different processes, are inherently linked through the nature of their processes (data processing and programming).

The specification requirements of Phase I were mapped to a solution, and an XML database was chosen to house the raw linguistic resources and associated localised setting data. An XML Document Type Definition (DTD) was drawn up which described how each of the datatypes would be structured and linked to the data around it. Once developed and complete, the database was populated with the relevant content.

4.6 Phase IV: Database proof of concept

The development of the proof of concept outputs from the linguistic content database began with an examination of the ways which the content could be plugged in to the product/display mechanism to illustrate how the content can be accessed and also employed in alternative types of resource outputs.

The initial departure point for this research had been the easier and faster creation of localised resources for teachers. As such, template- and paper-based handouts were one of the output types chosen. The second type of output illustrates how the content can be accessed and manipulated to form mini-lexica from the linguistic content database. The lexica are relevant to any one section of the content. The third output allows the teacher to create a content report which can be used for planning. HTML, XML, PHP and DTD were used to structure the database, code the teacher/practitioner interface and develop the outputs.

4.7 Summary and conclusion

This chapter drew discussions from the previous chapters together and added in the local context of the primary school modern language classroom in Ireland, to describe the combined global and local research space. It initially examined the resource challenges of the local context with its complex language setting, lack of coordinated and integrated language teaching and learning, and teachers' overuse of textbooks in place of the curriculum for planning, teaching and standardising their teaching content. The development of one instantiation of a localised linguistic content database was proposed as the solution to the global and local resourcing

challenges. Five localised settings were mapped out to fulfil the requirements of localisation of resources for the research context. The notion of localised setting (LS) was introduced as one means to localise resources to a specific context, as well as a means to reuse content where localised settings can be switched on or off in contexts where they are relevant or where they are irrelevant.

The research design overview was explored, along with the three-pronged multi-disciplinary methodological choices which support the four phases of work. The three overarching methodologies were (1) Design-based Research, (2) CL/NLP study and (3) Development work. These methodologies supported the development of the prototype linguistic content database as well as the analysis and characterisation of the raw linguistic resources.

The methodologies described in Phase I supported the pre-classroom preparation including raw linguistic content drafting and LS note drafting, the curriculum design, the classroom research and the database specification. The participants were also described in this section, along with the data gathering instruments used in the classroom.

Phase II described the CL/NLP and SLA basis for the metrics used to analyse and characterise the raw linguistic content, as well as to compare and contrast it to two other data sets. Phase III outlined the processes and technologies involved in developing and populating the linguistic content database. Phase IV described how the linguistic content database was used as the content source for a number of proof of concept outputs.

The next chapter describes the data which was gathered in the classroom through Phase I of work, and identifies the extent to which the drafted raw linguistic resources and associated LS notes were relevant and appropriate to the Irish primary modern language classroom.

Chapter 5 Localising and testing language learning content

The previous chapter outlined the processes involved in the development of localised content and associated localised settings (LSs) which account for the particular challenges and learning needs of the specific primary modern language learning context in Ireland. The aim of this chapter is to examine whether these processes were successful in localising the language learning content and LSs to their intended primary school audience. It examines whether the created language learning content and LSs presented through resources accounted for the children's localised learning needs.

Through this investigation, the evaluation of the raw linguistic resources and associated LS notes becomes the inverse of the original objectives outlined. Hubbard (1996) supported this method of evaluation, in a learner or teacher 'fit' value being applied for each of the original guiding needs and factors which were considered in the design space. If the learner fit is positive, the development process was successful in its aims. Chapelle (2001) also posited learner fit as one of six factors which should be considered in her CALL evaluation model. She described learner fit as "the amount of opportunity for engagement with language under appropriate conditions given learner characteristics" (Chapelle, 2001, p. 55). While Chapelle (2001) offers an appropriate evaluation scheme which considers learner fit, Hubbard's (1996) concept of learner fit better suits the context of evaluating content appropriateness, completeness and relevance in this context.

The learner fit in this research instance is analysed through ascertaining whether the LS notes were appropriate, whether the raw linguistic content presented is complete (is there any lacking or superfluous language) and how relevant the content is to its learners. This chapter also accounts for the additional content which was gathered (e.g. learner error analysis) in the local research context and which further localises the content to its context.

Section 5.1 describes how the data was analysed. Section 5.2 examines the data gathered across each of the three LSs which were examined in the local context – linkage and integration, Language Awareness and learner error analysis. Section 5.2 describes the data gathered on raw linguistic content and whether the content fit the local context for which it was developed.

5.1 Data analysis

The instruments used to gather data during the classroom-based research were introduced in Chapter 4 and can be found in the appendices (Appendix G and H). An Excel database was created to house the data as it was gathered in the classroom. Each of the data gathering instruments and the data itself was coded, so that relevant content could be transferred to the Excel database for analysis. The narrative text from open-ended responses was entered verbatim.

A new worksheet was created for each year in the classroom and for each class group. Data relating to different areas of work (e.g. with a focus on language awareness or error analysis) was all colour coded for ease of reference. All worksheets were combined into one file for final data analysis, but retained their class group and year as identifiers.

The data gathered across the instruments represented a mixture of open-ended fields and 2-and 3-point Likert scales. The use of Likert-scales allowed the data to be recorded and accessed faster through its ordinal form rather than using open-ended response types throughout. The error domains and categories were recorded by type and frequency. A numerical value was inserted under the relevant error header within the excel file to indicate the frequency of encounter of any one error type following observations made during any one lesson. Excel facilitated the analysis of data through filtered searches and in-built functions such as “countif”. Narrative data was coded where relevant or accounted for on an individual basis where too much variation existed between data items.

The focus of the data gathering in all cases was on the language, topics and resources presented (content), as well as the errors produced by the children. While this data was encountered through the course of teaching and learning activities, the activities were not the focus of observation. In each case within the data presented below, percentages are rounded to the nearest integer (and may therefore sum to more or less than the total indicated).

A total of 138 German lessons amounting to 103.5 teaching hours were held in the target space over a period of two and a half years (nearly three full school years). Lessons were only missed during this period due to school closures/holidays, class teacher illness or researcher unavailability (due to illness, conference attendance or research meetings).

5.1.1 Defining learner fit through localisation

As outlined in the previous chapter, the value of ‘learner fit’ was applied to an evaluation process within Hubbard’s model (1996). Hubbard’s evaluation phase was the inverse of the design phase

and involved checking in that all of the design criteria were met in the final product. The definition of 'fit' used in this research is the fit of *localising* the content to its learners. If the localisation process was successful, then the localisation fit is positive.

The instruments were designed to gather data which would help to gauge the level of relevance of the raw linguistic content and associated localised setting notes to the children in the classroom and to the primary school context in Ireland. Results will be discussed in the coming sections.

5.2 Evaluating raw linguistic content

The examination of *content* within each session allowed feedback to be gathered on the appropriateness of the raw linguistic content to the children in the session, any difficulties encountered with the raw linguistic content (e.g. pronunciation) and note to be taken of any additional raw linguistic content introduced or superfluous raw linguistic content presented. The examination of *resources/presentation* allowed data to be gathered on the clarity and age-appropriateness of the raw linguistic content, as well as any changes that may be required. The findings from *resources/presentation* are relevant to Chapter 7 but are not mutually exclusive from the discussion on raw linguistic content. A final more open-ended field allowed for any ancillary researcher's notes to be recorded.

The structured feedback forms (see Appendix G) were completed, where possible, as the classroom sessions proceeded. In all other instances, the feedback forms were completed within one hour of leaving the school. The feedback forms were then transcribed into the Excel spreadsheet within the same week they were initially recorded.

Raw linguistic content was not presented in the same order it is outlined in Appendix E. Content was selected according to its suitability in being the next relevant topic after any other topic(s) had been covered. Equally, the participants' input sometimes dictated the next relevant topic to be covered. All sections were explored with Class Group R, while only a few were explored during the shorter period available to work with Class Group S. This reflects the 'pick and choose' ethos which the raw linguistic content fosters in its linguistic content database form.

5.2.1 Focus on content

Testing the raw linguistic content went hand in hand with ensuring that the target space reflected one for which the raw linguistic content had been developed. For this reason, teaching resources were not limited to only those which comprised the raw linguistic content or were developed for

the specific purposes of this research. As mentioned above, it was important to mimic the context teaching space of an MLPSI school in as far as possible. This meant that the resources used in the classroom were always initially those which had been developed from the raw linguistic content. To further develop the relevant topics and raw linguistic content, further tailored resources were created, or alternative teaching activities or resources were employed. 22% of resources used in the classroom were supplementary to those developed from the raw linguistic resources (see Figure 5.1 below). For example, initially Handout 2 and Worksheet 2 (see Appendix F) were used to introduce the parts of the body. This raw linguistic content was then further examined through the total physical response (TPR) activity – Pumpernickel sagt (Simon says). Poetry and song were also used to reinforce the raw linguistic content.

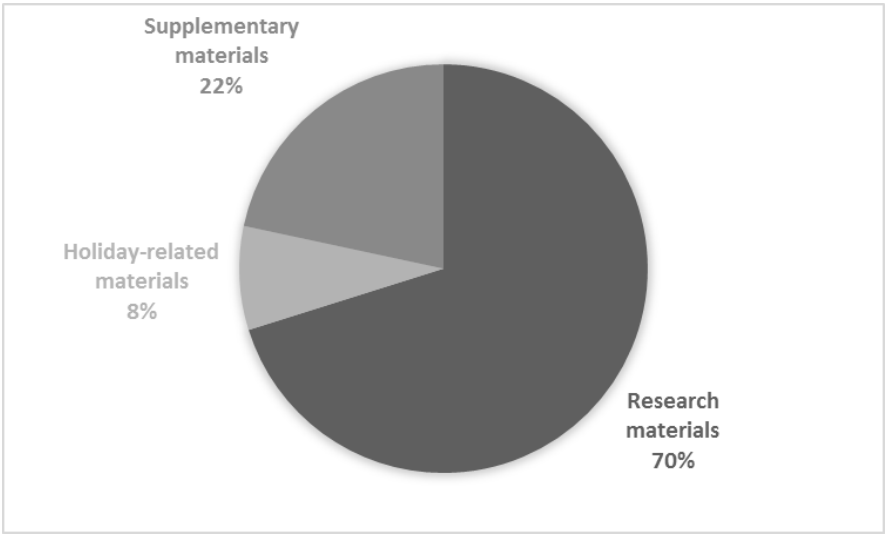


Figure 5.1 Types of resources used in the classroom

The supplementary materials used included poems (29%), hard-copy materials such as dictionaries, posters and flashcards (24%), children's German language learning book materials (19%), songs (14%) and a European Language Portfolio (ELP) (10%). The holiday-related materials were developed specifically for the children in the classroom but did not account for the content of the raw linguistic resources.

In keeping with the three resource-types used in the classroom, Table 5.1 outlines the proportions of additional words which were noted as being "relevant language which is missing". This was content which would have been required, encountered and/or requested during the course of teaching raw linguistic content topics. Some of this content was adopted into the end linguistic content database where it fit within the specified topics. Some of the additional content was already covered off in an alternative section (e.g. alt / *old* is in Section Nine on *Family 2 – Describing yourself and your family*) and so was not included a second time. Other content, which had been identified through poetry or song for example, was not integrated as its makeup

did not reflect the curriculum topics, but were words encountered through the course of language learning in the classroom.

	Relevant language missing from the raw linguistic content(number of words, n)	**Missing language as a percentage of total words of raw linguistic content v1.0 (n=2,263)**
Research materials	209	9%
Supplementary materials	129	6%
Holiday-related materials	34	2%
Total	**372**	**16%**

Table 5.1 Relevant content noted as 'missing' during lessons

Some examples of the missing language are summarised alongside their category of presentation (research materials, supplementary materials or holiday-related materials) in Table 5.2 below. Additions to language noted through research materials tended to occur as a result of language which emerged through reinforcing the language taught by using it in new ways. Table 5.2 below provides an example of this, with the addition of phrasing to ask for a telephone number and how much pocket-money one gets within Section Three on *Numbers.* Content identified through supplementary materials such as poetry or song tended to be more difficult to match to any of the sections. While some of their content was relevant to section topics (e.g. *Tageswochen* by Erich Weinert / *days of the week* which also covered weather), the relevant language would already have been included in the initial draft (v1.0) of the raw linguistic resources. Additional content which was recorded in any one section, did not necessarily fit within that section and if carried through for use within the linguistic content database, may have been inserted into an alternative, more relevant section. The language noted around the holiday-related materials was all relevant to celebrating the season and integrating with the classroom learning and school environment. Similarly, it was an opportunity for cultural awareness activities in comparing and contrasting traditions from Ireland and German-speaking countries. However, cultural elements were not included in the raw linguistic content and did not carry through to the linguistic content database unless they were relevant to other sections (e.g. Der Stern / *the star* carried through to Section 18d on shapes).

	Section topic being taught	Sample language noted as 'missing'
Research materials	Section 3: Numbers 1-20	Welche Telefonnummer hast du? / *What is your telephone number?* Meine Telefonnummer ist X / *My phone number is X* Wieveil Taschengeld bekommst du? / *How much pocket money do you get?* Ich bekomme €5 Taschengeld pro Woche / *I get €5 per week*
	Section 17: Food	Ich hasse / *I hate* Ich mag / *I like* essen / *to eat;* trinken / *to drink* Noch etwas? / *Anything else* Das ist billig/teuer / *that is cheap/expensive*
	Section 18e: Colours	Das Auto ist X / *The car is X* Der Schneemann ist X / *The snowman is X* Die Wolke ist X / *The cloud is X* Die Sonne ist X / *The sun is X* Die Blume ist X / *The flower is X*
Supplementary materials	Section 1: Greetings ▪ ELP ▪ Children's German language learning book	England / *England* Afrika / *Africa* Frankreich / *France* Spanien / *Spain*
	Section 8: Weather ▪ Poem: Wochentage (von Erich Weinert) ▪ Poem: Regenwetter[15] (von Friedrich Halm)	Der Himmel / *the sky* Die Plätze / *(all over) the place* Der Sperling / *the sparrow* Das Dach / *the roof*
	Section 17: Food ▪ Poem: Beim Kaufmann (von James Krüss) ▪ Dictionary work	Die Milch / *the milk* Der Krebs / *the crab* Der Teebeutel / *the tea bag* Die Pizza / *the pizza* Der Hamburger / *the hamburger*
Holiday-related materials	Lent / pancake Tuesday	Das Mehl / *the flour* Das Ei, / *the egg* Der Teig / *the batter* Die Pfannkuchen / *pancakes* Eier aufschlagen / *to crack eggs*
	Easter	Frohe Ostern / *Happy Easter* Der Osterhase / *Easterbunny* Das Osterei / *the Easter-egg* Die Osterlocke / *The daffodil* Das Ostern / *Easter*
	Christmas	Der Engel / *the angel* Die Kerze / *the candle* Die Krippe / *the crib* Das Weihnachten / *Christmas* Die Weihnachtskarten / *the Christmas cards*

Table 5.2 Examples of relevant content noted as 'missing' by resource type and section

The next area of focus on the structured feedback form related to "superfluous or irrelevant language" within the raw linguistic content. The challenge with identifying superfluous language

[15] Poem sourced from Project Gutenberg DE http://gutenberg.spiegel.de/buch/gedichte-6586/2

during a lesson was that content from any one section was sometimes covered over one or more lessons. As a result, it was difficult to limit the boundaries of raw linguistic content to any one lesson for recording the superfluous language. Some of the content was superfluous to what was required for one class group, but as the content within the content database is selected and customised by the teacher, this raw linguistic content was retained to provide greater variety among the language available for selection, and to ensure that the raw linguistic content available could cater to a wider range of children (e.g. in urban/rural locations, male/female).

The third focus question on the structured feedback form queried how relevant the content was to the children in the classes. A three-point Likert scale was used to note answers, with additional space for any further comments. This question was included to ascertain whether the children would need and use the language of the raw linguistic content to talk about themselves, their lives and environment, which represent the first and second sphere of relevance to children learning languages (McKay, 2006). In 100% (n=119) of cases in the data gathered, content was deemed to be "very relevant" to the children in the class (Table 5.3). This could be attributed to the fact that the curriculum-bounded content was specifically geared towards children of this age-group and in this Irish setting. Similarly, the raw linguistic content developed around the modern language curriculum documents was chosen to reflect the children's age, location and learning environment.

Materials		**Count**			**Count**			**Count (n)**	
		n	**%**		**n**	**%**		**n**	**%**
Relevance to children	Not relevant	0	0	Relevant	0	0	Very relevant	119	100
Difficulties encountered	Some difficulties encountered	11	9				No difficulties encountered	109	91

Table 5.3 Results of data gathering with a focus on content

The final question within the structured feedback form which focused on content examined whether any "difficulties" were encountered with the content presented. Some 9% (n=11) of entries for this area of focus indicated that difficulties had been encountered. The types of difficulties outlined (N=11) (see Appendix I: 4a) included

1. children were having problems grasping the order reversal of numbers beyond 21 (directly translated as 1 and 20, 2 and 20 etc.) and that this challenge needed to be resolved before dates and family members' ages, for example, could be adequately covered (27%, n=3)
2. some of the research materials and supplementary materials contained too much new vocabulary to try to work through in one to two lessons (e.g. Handout 6 – Was ist das / *What is that?*) (36%, n=4)

3. some of the supplementary resources were a little too far beyond the competency level of the participants for them to be able to traverse them, even with the assistance of accompanying glossaries (18%, n=2).

This third finding, although indicated through a low level of data, highlighted the importance and need for localised and customised resources, as ready-made supplementary resources did not represent a perfect match to the practitioners' and/or learners' needs or preferences (although a close match as they were selected on grounds of their suitability). The second and third points signify a need for an awareness of the level of complexity of a text for learners – Decoo (2011) cited Waring and Nation (2004) who assert that "if the text is too difficult then little learning can take place – especially if the known coverage rate is lower than 98%" (p. 111). Decoo (2011) further described how "the *threshold hypothesis*…asserts that processing strategies cannot function in L2 if language-specific knowledge, in particular vocabulary, is below a certain threshold or ceiling" (p. 112). Furthermore, findings from Decoo (2011) and Waring and Nation (2004) asserted that learners will learn "3-6 words per hour of reading" (Waring and Nation, 2004, p. 106) and that words from the same semantic field should be limited to avoid confusion (Decoo and Colpaert, 1997).

5.2.2 Focus on resources and presentation

The structured feedback form also contained three areas of focus which centred on the clarity and appropriateness of the resources and their presentation. While there was some overlap between the content and its presentation mode, the results of this section are equally relevant to this chapter and Chapter 7 on developing resource outputs from the linguistic content database.

The structured feedback form presented a three-point Likert scale for recording data relating to the clarity of the resources used. 98% (n=117) of resources were deemed to be "very clear" and 3% "clear" (see Table 5.4). None of the resources were recorded as being "not clear". Of the 3% (n=3) (see Appendix I: 5a), two entries referred to supplementary resources. Just two notes were recorded alongside this data. The first again pointed to the challenge of the numbering system. The second pertained to children working through project information on European countries and the languages spoken in them (to coincide with the European Day of Languages on 26 September). Each child was given country information printed from an Encarta Encyclopedia CD (in English) through which they needed to work to find their relevant country information. Some challenges were experienced with English literacy and the children's ability to pinpoint the relevant content.

Materials		**Count**			**Count**			**Count (n)**	
		n	**%**		**n**	**%**		**n**	**%**
Clarity	Not clear	0	0	Clear	3	3	Very clear	117	98
Age-appropriate-ness	Not age-appropriate	0	0	Age-appropriate	3	3	Very age-appropriate	116	97
Changes required	Change required	10	8				No change required	110	92

Table 5.4 Results of data gathering with a focus on resources and presentation

97% (n=116) of resources were indicated to be "very age appropriate" and 3% (n=3) "age appropriate". All of the comments associated with the "age-appropriate" resources related to supplementary resources (see Appendix 1: 6a). One of these entries referred to a BBC Deutschland Plus 1 episode which the children had watched and which introduced vocabulary which crossed over with Section 1 of the raw linguistic resources. While a beginner's series, with language which had been learned and practised, the focus of the storyline was on a Gastarbeiter (guest worker) in Germany and was one with which the children could not personally relate.

92% (n=110) of resources were noted as not needing any editing while 8% (n=10) were noted to require editing. The comments on the latter entries spread evenly across the three resource types in use and returned to those points made above under the discussion on difficulties encountered (see Appendix I: 7a). Suggested changes included breaking down content across multiple handouts or lessons (n=6) and the use of more images to help clarify the content (n=1) (e.g. The handout on Fastnachtsdienstag / *Pancake Tuesday* in Appendix F). These points reinforce those made above about the threshold hypothesis and children's levels of understanding.

The 'other notes' (n=46) (see Appendix I: 8) which were gathered pertained to successes and challenges in the target space, as well as those relating to the raw linguistic content and resources in use. Successes reflected content that was effective, activities which were enjoyed by participants, the reinforcement and building of content within one activity over time (e.g. our news which gradually progressed to include details about themselves, the weather, the colours associated with the weather/outside environment etc.) and creating an element of competition among the participants. Challenges outlined related to the need for much revision and reminding in the target space, the requirement for much additional explanation of linguistic/grammar foundations (e.g. what is a noun) and the requirement to present fewer new nouns in one sitting. Other comments related to the resource in use and the teaching and learning which was taking place.

5.3 Gathering evidence of the localised settings

A total of five localised settings were identified for the primary modern language classroom in Ireland. Three localised settings were relevant to the classroom-based research. The PSC and modern language curriculum documents were essential in preparing the raw linguistic resources and in preparing to act as a teacher in the classroom, but no specific data was gathered on these localised settings. The notes on relevant areas of integration and linkage and Language Awareness were investigated in the classroom for their appropriateness to the children and the learning environment in the primary school classroom and school. The final localised setting, learner error analysis, required specific data to be gathered. This latter data was gathered according to the closed-ended checklist described in the previous chapter. Entries were noted by type and frequency of occurrence. The following sections describe data gathered on these three localised settings.

5.3.1 Linkage and integration topics from the PSC

The classroom-based focus on the integration and linkage notes concentrated on their appropriateness to the setting (German language learning with a visiting teacher) and the children in the classroom, as well as teaching in the surrounding environment of the classroom (the PSC with the class teacher) and in the school (for school-based activities and focuses).

As outlined in Chapter 4, integration notes which were identified according to raw linguistic content sections were too extensive to allow for each point to be examined in the classroom. However, as time in any educational setting is limited, the insufficient time for testing the full integration and linkage topics highlighted the broader system issues, including *lack of time*, which existed around the integration of modern languages at primary level within the MLPSI (Harris and O'Leary, 2012; NCCA, 2008). Further system issues related to analysing who was best placed to teach the modern language (the classroom teacher or a visiting specialist teacher) with consideration being given to integrating other subjects of the PSC and making time for the modern language (Harris and Conway, 2002; Harris and O'Leary, 2012; NCCA, 2008). Classroom teachers were cited to be better placed to integrate the modern language with other subjects of the PSC (Harris and O'Leary, 2012). In instances where visiting teachers facilitated the modern language, research showed that it was easier to block off time for the modern language and ensure teaching took place each week (Harris and O'Leary, 2012).

The most relevant areas of integration with the PSC were selected according to their relevance to the teaching and learning going on in the classroom outside of the German class, and around

the school (e.g. holidays and festivities). For instance, much teaching time was devoted to relevant holidays - Irish ones, German ones or those which were celebrated in both locations.[16] While the raw linguistic content did not include holiday content, 8% of lessons (holiday-related materials) were attributed to them. Those integration topics which were relevant to the raw linguistic content and which were used in the classroom, were recorded as being appropriate to the children and their learning.

The notes on linkage referred to other sections within the raw linguistic content where reinforcement could be accounted for (through the content in a previous section), or additional relevant content could be drawn in from a later section. The majority of linkage topics covered in the classroom were the former type, calling on previously covered relevant content. 37% (n=54) of lessons included content from more than one section from the raw linguistic content. The majority of these lessons referred to earlier-covered sections for revision/reinforcement-purposes or progressed the learning by moving ahead with related, but new, content. The need for frequent revision in the classroom could have been symptomatic of some of the challenges associated with disadvantaged settings, such as the need for additional revision and little homework being completed. Of note, is that the proportion of multi-section lessons linking between sections for Class S was greater from the outset of their German lessons, than for Class R, as illustrated in Table 5.5 below. Although the number of lessons held for Class S is substantially fewer than for Class R, the proportion of multi-section lessons is far greater. This could have been attributed to my own changing selection of materials and revision schedule having learned what worked best with Class R (greater levels of revision and topic linkage), or to a faster pace of progression with Class S than Class R.

	Class R		**Class S**	
	n	**%**	**n**	**%**
Single-section lessons	83	67%	6	33%
Multi-section lessons	40	33%	12	67%
TOTAL	**123**	**100%**	**18**	**100%**

Table 5.5 Proportions of single- and multi-section lessons for Class R and Class S

5.3.2 Language Awareness for Gaeilge/Irish, English and German

Examination of language awareness notes in the classroom focused on the similarities and differences across the grammatical and lexical features of the three languages – English, Irish

[16] While 'German' holidays are the focus of discussion, this refers to celebrations which take place in German-speaking countries.

and German. The Language Awareness notes which were developed (see Chapter 4) to sit alongside the raw linguistic content were sometimes repeated across sections but referred to the vocabulary of the section within which it sat. For example, the language awareness notes in all sections referred to new German words (presented through the section) which were similar to their English and Irish equivalents.

Reflective cues were used to note observations about the appropriateness of language awareness notes. The results of the reflections gathered indicated that in 100% of cases, the notes which had been prepared were appropriate for the primary school children. All of the content was relevant to the children, as it reflected the learning across German, which integrated and overlapped with known language in English and Irish, or language being learned in Irish. The listing of language awareness notes per section was not always completely covered, but the same themes emerging across sections ensured that each one was addressed in at least one section. Challenges arose in connection with the frequency with which covered language awareness themes were required to be revised, rather than any inappropriateness or shortage of relevant themes being noted.

One of the challenges which arose around the inclusion of language awareness was the necessity to continually revise some aspects of the grammatical labelling used in linguistics and language learning, for example, what the parts of speech were and how to identify them. Children were able to rhyme off the 'definition' of parts of speech, for example, '*a noun is a person, place or thing*' which had been learned with their classroom teacher, but found it harder to identify the noun(s) within a sentence. This led to many challenges with the capitalisation of German nouns, or the incorrect capitalisation of other parts of speech, which was evidenced through the 'incorrect case' error being the most frequently noted error (20% of total errors - see Section 5.3.3 below).

Another challenge centred on genders being assigned to nouns, and the requirement to learn a noun's gender along with its form/spelling. While Irish does have masculine and feminine nouns, noun gender is rarely ever explicitly taught in primary English-medium schools. The third highest error type encountered was a lack of article-noun agreement (11% of total errors recorded).

A key benefit to focusing on grammatical areas of language awareness was the reinforcement of basic and core areas of relevance to all language learning in the sphere of the children. Thouësny (2011) noted that "[students] may have struggled with terms describing specific aspects of the language, such as auxiliary, direct object, indirect object, or pronoun. In fact, students who come through the Irish educational system tend to have a deficit in terms of grammatical knowledge"

(p. 246). While some of the parts of speech referred to by Thouësny (2011) would have been beyond the scope of the primary school children, it was important to begin to form the foundations of an awareness of language and its makeup.

A further examination of the student errors (see the next section below) which emerged revealed that some additional areas of Language Awareness were very relevant in the classroom and learning German. These included code switching with English (7% of errors), use of diacritics/accents (5% of errors), and word order (4% of errors). The full listing of errors encountered and recorded according to frequency are listed in the next section.

5.3.3 Learner error analysis

The examination of learner errors in the classroom was a gathering exercise, rather than a checking one, which was the case with the raw linguistic content and other localised settings. The learner error data transferred to the Excel spreadsheet indicated the number of occurrences of each error made per lesson, and from written work and during the oral test. Cross-tabulations revealed the sections which were being taught at any stage for the rows of data entered for error analysis. A total of 7,075 errors were recorded across the two and a half year period in the classroom. The total number of errors recorded during and/or just after *any one lesson* ranged from a minimum of 2 errors (e.g. 2 December 2003, Section 3: Numbers 1-20) to a maximum of 163 errors (17 November 2005, Section 13: Pastimes and hobbies), reflecting the progression from easier to more challenging topics and language. Error tags used were not nested as individual tags, rather they are concatenated with the under-score used to separate the constituent parts of the single error tag. For example, error <W><FM><CA> was recorded as <W_FM_CA>.

The overall results of the error analysis revealed the most common errors made by domain (Figure 5.2). A more detailed listing of the frequency of occurrence of each individual error type is recorded in Appendix K with the top ten of these outlined in Figure 5.3 below. The written format of children's work was the source of the most errors recorded (37%) during the time in the classroom. The most frequently encountered error of all occurred in this domain and was the use of the incorrect case (<W_FM_CA>) (Figure 5.3). This error type accounted for 20% of errors made. As all nouns in German receive a capital letter, it was a challenge for children to first identify the nouns in a sentence, and then to assign it a capital letter. In some cases, children applied a capital letter to a mixture of parts of speech, where they could not identify the nouns (e.g. see Figure 5.5 below)

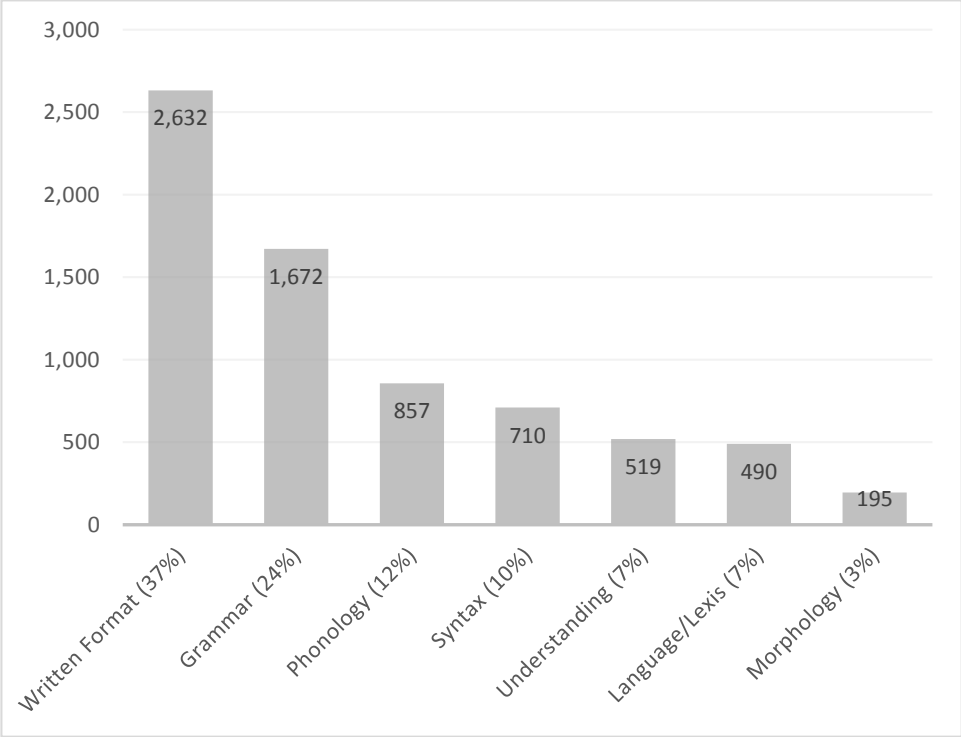


Figure 5.2 Total errors recorded during classroom research (by domain) (N=7,075)

Grammar was the domain receiving the second highest total count of 24% of errors made, and phonology the third highest at 12%. Difficulties with pronunciation (<P_Pron>) was the second highest error type recorded overall, with the full 12% of errors from the domain being attributed to this single error type (Figure 5.3). The third most frequent error type recorded was article-noun agreement (<G_AT_ArtN>) with a total of 11% of errors made. German has three noun genders where the article and noun need to match. This is a new concept for native English speakers, although Irish does have masculine and feminine nouns (as mentioned above, it is unfortunate that this distinction is rarely made in teaching). The remaining error types within the top ten encountered are outlined in Figure 5.3 below. Together they made up 85% of the errors recorded.

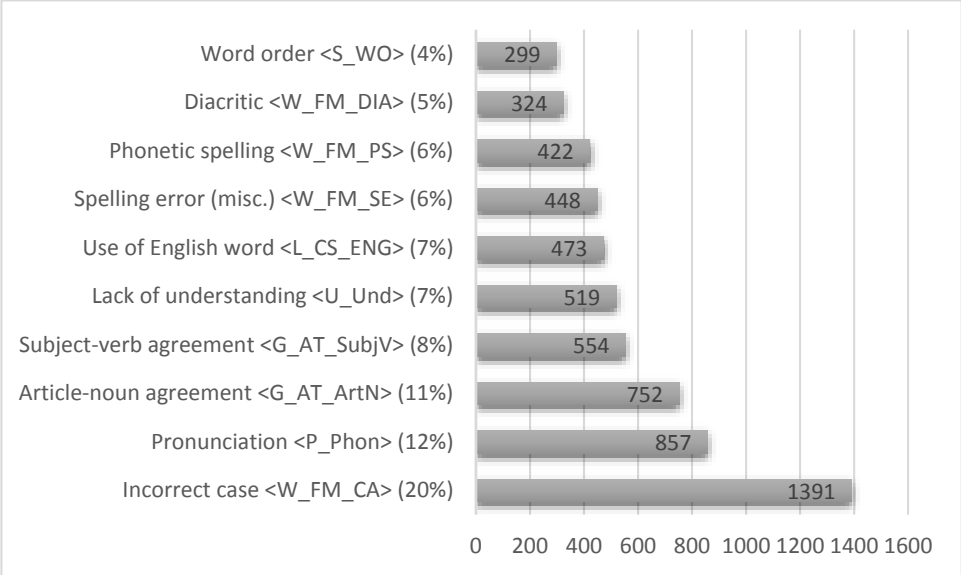


Figure 5.3 Ten most frequent error types encountered in the classroom

Aside from classroom observations, the oral test and written work also accounted for the error totals outlined. The oral test was carried out with Class R during the second year of work in the classroom. Table 5.6 below provides an example of a transcribed oral test. While the error domains and categories were used to account for the typical errors made by the participants, they did not account for the stilted answers received in some cases or a lack of language competence in other cases. Table 5.6 below provides one example of an instance where the child receives much scaffolding and assistance from the teacher/researcher to facilitate him in providing an answer to the prompts. Errors noted in the classroom and from the oral test were both classified according to type and frequency. Error tags are inserted below and highlighted for clarity. Lack of understanding (<U_UND>) was also noted to allow for revision and assistance where needed.

File #13		
Time	**Speaker**	**Narrative**
.00	Katrina	Ok, so this is *Student X*. And your first question. Wie sagt man 'hello' auf Deutsch? / *Ok, so this is Student X. And your first question, How do you say 'hello' in German?*
.05	*Student X*	Hallo / *Hello*
.06	Katrina	Hallo, gut, Und wie geht's? / *Hello, good, and how are you?*
.09	Student X	Ich bin [?] */ I am [?]*
.12	Katrina	Bist du gut, oder schlecht / *Are you good or bad?*
.14	Student X	Gut / *good*
.15	Katrina	Gut ok, und wie heißt du? / *good, ok, and what is your name?*
.18	Student X	<U_UND>Ich bin zehn Jahre alt</U_UND> */ I am ten years old*
.22	Katrina	Ok, und wie heißt du? Heißt du Dean, oder heißt du Gareth? Oder heißt du Frau Keogh? / *Ok, and what is your name? Is your name Dean, or Gareth? Or is your name Ms. Keogh?*
.30	Student X	Oh, …
.31	Katrina	You say, ich … heiße… / *You say, my name is …*
.33	Student X	Ich heiße Student X / *My name is Student X*
.34	Katrina	Student X, gut. Und woher kommst du Student X? / *Student X, good, and where do you come from Student X?*
.38	Student X	<U_Und>[pause]</U_UND>
.41	Katrina	Du kommst aus Frankreich, oder Deutschland, oder Irland? / *You come from France, or Germany, or Ireland?*
.47	Student X	<U_UND>Oh</U_UND>
.48	Katrina	Ich komme… */ I come …*
.50	Student X	Ich komme aus <L_CS_ENG> Ireland</L_CS_ENG>/ *I come from Ireland*
.52	Katrina	Irland? / *Ireland?*
.53	Student X	Yeah
.53	Katrina	Ok, gut. Und wo wohnst du / *Ok, good, and where do you live?*
.56	Student X	<U_UND>W… What?</U_UND>
.57	Katrina	Wo wohnst du in Irland? Where do you live in Ireland?
.58	Student X	[breath]
1.01	Katrina	Ich wohne… / *I live in*
1.04	Student X	Ich wohne <S_MW>[-]</S_MW> place X */ I live [-] place X*
1.05	Katrina	*<u>In</u>* place X, gut. Und hast du Geschwister? Hast du einen Bruder oder eine Schwester? */ <u>In</u> place X, good. And do you have any siblings? Do you have a brother or a sister?*
1.13	Student X	[pause] Ich eine <P_PRON> Wester</P_PRON> */ I a sister*
1.16	Katrina	Ich habe eine Schwester, gut. Ok, und wann hast du Geburtstag*?./ I have a sister, good, ok, and when is your birthday?*

Table 5.6 Excerpt from oral test transcription
(note: Student X and Place X were inserted as replacements for student information to maintain anonymity)

A total of 43 errors were noted across the transcriptions of the oral tests of 13 students. The most frequent errors recorded for the oral test were use of an English word in place of a German one (<L_CS_ENG>) (n=7), article-noun agreement (<G_AT_ArtN>) (n=7) and the insertion of an additional word (<S_EW>) (n=6). Figure 5.4 provides a listing of the errors made in the oral test by domain.

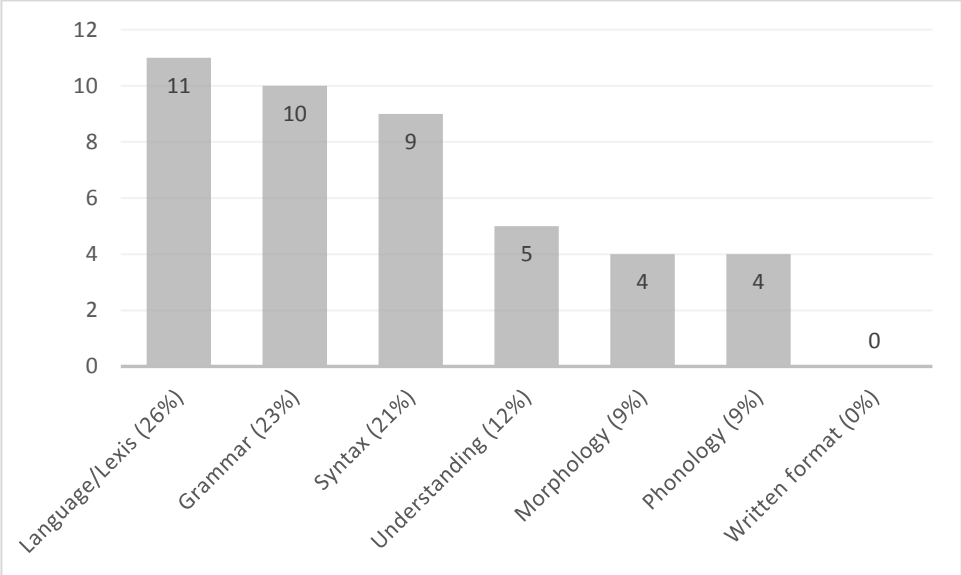


Figure 5.4 Oral language test errors by domain (N=43)

Students' **written errors** were also analysed by type and frequency. Figure 5.5 and Figure 5.6 below provide samples of the types of written work completed by children and the associated error tags which were recorded by type and frequency for each lesson and any follow-up homework activity. Error tags were inserted and used for research purposes, but were never used directly with the children.

Figure 5.5 Sample A of children's written work

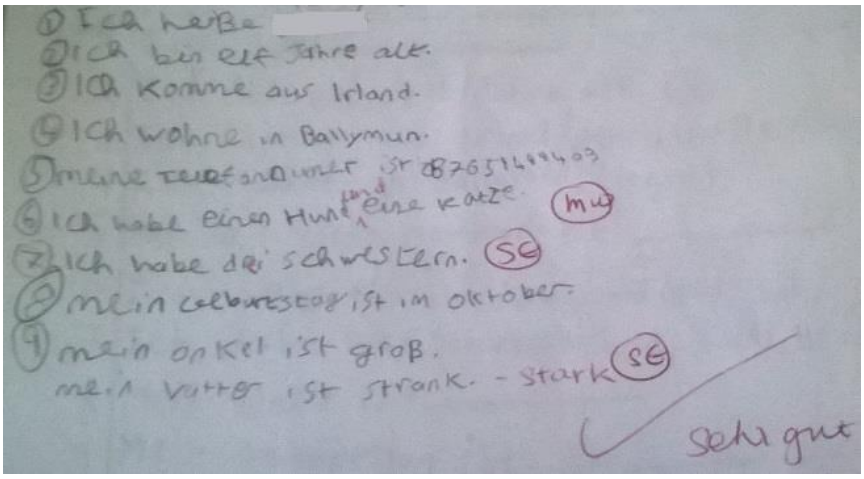


Figure 5.6 Sample B of children's written work

The typical errors recorded formed part of the content in the linguistic content database (see Chapter 7) and accounted for further localisation to the local context. Those encountered provided evidence and samples of the types of errors which can occur in settings similar to the research context. The errors encountered reinforced the language awareness topics which were outlined above and provided a guideline for practitioners on typical pressure points which need to be accounted for in teaching and learning German in settings similar to the research context. Learner errors helped to inform an extended version of the language awareness notes for the linguistic content database.

5.4 Gauging the level of localisation success (learner fit)

The results of the data gathering process on content revealed that the raw linguistic content was suited to its local context (the children in the classroom and the modern language curriculum documents) and that the development processes outlined in the last chapter were suitable for localising content to its local context.

While the localised settings were not fully tested in the classroom as their listing was too long to allow each topic to be trialled, those which were were shown to be suited to their local context, specifically - the children, the local PSC, teaching guided by the classroom teacher and the school environment.

The learner fit was deemed to be positive, with the localisation process a success for its local context.

Two questions are raised with the level of positive data recorded for the raw linguistic content and associated localised settings. (1) As an individual researcher, who was both teacher and researcher in the classroom, was enough scope and time provided to adequately observe and track the areas of investigation? (2) As content developer and data gatherer, was enough neutrality maintained in data observation and recording? Conversely, it could be argued that the content was developed and localised specifically to its context curriculum and learners, and hence, that the results could equally reinforce the benefit of the development process outlined in the last chapter. This potential bias could be readdressed in future work where the research design could ensure a greater distance between the content author and local context researcher.

Aside from accounting for learner fit within this chapter, the classroom-based work also helped to account for the file types and relevant content which transferred to the linguistic content database. Analysing the data gathered helped to reveal how each of the data types were best associated with one another, and how they could be used to assist in the teaching and learning of a modern language in the local context. For instance, typical learner errors aligned well with language awareness notes. Learner error analysis helped to identify challenges which can emerge for children in a learning context similar to the local research context.

5.5 Summary and conclusion

Section 5.1 described how the data gathered from each of the areas of evaluation and focus was coded and transferred to Excel format. It outlined how the data was analysed through filtered searches and examinations of the data, recoding the data to present it according to theme and summing the relevant filtered data searches.

Testing the enhanced raw linguistic content (Version 1.0 with LS notes) in the local context emphasised the pushes and pulls of working in a local context which requires a realistic contribution (as teacher) and mutual benefit for the participants and researcher. While 78% of the content used and tested was developed specifically for the target space from the raw linguistic content, the remaining 22% was required to maximise the learning benefit of the participants (alternative activities/genres/resources for reinforcing raw linguistic content initially presented). The deviation reinforced the notion that teacher-selected supplementary materials will always require some element of localisation, with all of these resources being noted as requiring customisation for the participants. All of the challenges noted with the resources (full 9% of resources with which difficulties were encountered and the full 8% of the resources requiring changes) were attributed to supplementary resources.

Section 5.3 outlined findings from evaluating the localised setting data – (2) Linkage and Integration and (4) Language Awareness. The notes generated on the relevant areas of linkage, integration and language awareness were more extensive than envisaged which resulted in just a few of the areas identified being investigated through the course of teaching. Those which were examined were relevant to the local context and to the raw linguistic content. Data gathered on language awareness illustrated the relevance of the topics identified. This was reinforced when the typical learner errors were analysed, and highlighted the areas requiring specific attention.

The data gathered on the frequency and type of learner errors was also described, with examples provided from the oral and written evidence gathered. The most frequent type of learner error noted was in written format, with 37% of errors being attributed to this domain. Grammar (24%), phonology (12%) and syntax (10%) represented the next three highest occurrences of errors noted. The addition of data and information relating to these localised settings to the raw linguistic content increase the localised, tailored and focused nature of the raw linguistic content to its local context.

The next chapter progresses to an examination of Phase II of the research, and the results of the analysis and characterisation of the raw linguistic resources and other language learning resource data sets.

Chapter 6 Analysis and characterisation of raw linguistic content (v1.0)

This chapter describes Phase II of the research design which involves processing the language contained within the raw linguistic content and analysing its makeup so that it can be characterised and described. The analysis also reveals the degree to which the raw linguistic content exhibits features of a controlled language.

The processing of the raw linguistic content facilitates the addition of part of speech (POS) tags which allows the identification and isolation of certain parts of speech relevant to the analysis carried out. Section 6.1 describes how the raw linguistic content was processed to add part-of-speech tags.

Three datasets are introduced for use in the analysis and characterisation of language learning resources. The first is the content of v1.0 of the raw linguistic resources. The second dataset is a series of samples taken from the TIGER corpus of German newspaper language. The third is a sample taken from the content of a beginner's adult German language learning textbook. Section 6.2 describes the three datasets, the matrices used (from the SLA and CL/MT domains) to analyse and characterise them across two phases of testing and the results obtained.

6.1 Processing raw linguistic content

This phase of work presented an opportunity to process the raw linguistic content before the compilation of all authored, gathered and tested data in Phase III. Once this data is added, the content becomes more cumbersome for processing (and its format will also differ to the raw linguistic content in most instances).

As mentioned in Chapter 3, many CL/NLP tools are available which can be used to add annotations to raw linguistic content. Tools such as lemmatisers, machine translation tools and frequency counts and blocks can be used to process, inform and enhance linguistic content. The benefit of investigating these processes can be seen through the work at Didascalia/Linguapolis (Decoo and Colpaert, 1996; 1997; Decoo et al, 1996; Decoo, 2011) as discussed in Section 3.2.5 (p. 82). Processing in this research instance included pre-processing to change the file type and layout of the raw linguistic content, part of speech (POS) tagging and post-processing to make the raw linguistic content more accessible and readable for the analysis phase of Section 6.2. The coming sections describe how the raw linguistic content was processed and altered. The

purpose of this processing was to prepare the content of the raw linguistic resources for analysis, as well as enhancing the content with the addition of POS tags.

6.1.1 Pre-processing

The total running word count for the raw linguistic content was quite small at only 2,168 words in total (this is not the token number and excludes punctuation marks). Each section contained full sentences as well as relevant vocabulary items and partial sentences (for example, er ist … Jahre alt / *he is … years old*).

While MS Word format was used for the early stages of work to allow use of and access to the raw linguistic content in Phase I, it was unsuitable for processing the language content directly. The content was trimmed down to raw text for processing and formatting was removed. The raw linguistic content was then converted into a .txt file.

Each sentence and vocabulary item was moved to its own line and followed by a full-stop (see extract in Appendix L). An asterix was inserted before the names of subsection titles for easier identification. The full stop was used as a marker to indicate the end of the word/phrase/sentence boundary. The next stage was to use a part-of-speech tagger on the processed text for the addition of parts of speech to each token. The final token count after pre-processing was 2,887 tokens (see Table 6.8, p. 203).

6.1.2 Part-of-Speech tagging

Schmid's TreeTagger (Schmid, 1994) is a language independent part-of-speech (POS) tagger. It requires a lexicon and a manually tagged training corpus to allow it to be used on any one language. TreeTagger annotates texts with its parts of speech and lemma information. Lemmas are the base form of words, for example, the lemma of 'is' is 'be', and 'words' is 'word'. TreeTagger was run on the raw linguistic content text file. An extract of the POS-tagged file is provided below in Figure 6.1.

```
1.          ADV     1.
Greetings           NN      <unknown>
.           $.      .
Hallo       ITJ     hallo
!           $.      !
Guten       ADJA    gut
Morgen      NN      Morgen
.           $.      .
Guten       ADJA    gut
Tag         NN      Tag
.           $.      .
Guten       ADJA    gut
Abend       NN      Abend
.           $.      .
Gute        ADJA    gut
Nacht       NN      Nacht
.           $.      .
Schlaf      NN      Schlaf
gut.        NE      <unknown>
Auf         APPR    auf
Wiedersehen         NN      Wiedersehen
.           $.      .
Tschüs      NE      <unknown>
.           $.      .
Bis         KON     bis
bald        ADV     bald
.           $.      .
Bis         KON     bis
morgen      ADV     morgen
.           $.      .
Bis         APPR    bis
Donnerstag          NN      Donnerstag
.           $.      .
*           XY      *
```

Figure 6.1 Extract from the POS-tagged file of the raw linguistic content (prior to checking for tagging errors). Note: Line numbers are only present in the file viewer and not in the actual text file. Tagging errors are highlighted in red.

Once the file was tagged, it was manually examined for accuracy. Changes were made to the tags where incorrect POS tags had been assigned by the tagger. The corrected tagged text was not used to retrain the tagger as there was no further need to use TreeTagger on the data.

Schmidt (1994) reported 96.36% accuracy when TreeTagger was tested on a sample dataset from the Penn-TreeBank. The Penn-TreeBank is an annotated corpus which shows the syntactic structure of the language within it using trees. The raw linguistic content text file contained 3,393 tokens[17] (including punctuation marks, comments in English which describe the additional marks used (e.g. the * was used to precede Section heading names in English) and section and subsection headers). Excluding those errors made by the tagger when it encountered English (the German data was in use with the tagger), 170 errors were made. This gave an accuracy rate of ca. 95%. Factors which attributed to this slightly lower accuracy rate were:

- Partial sentences

[17] The previous section referred to 2,168 words in the raw linguistic resources. This figure does not account for punctuation which is an element which is tagged by TreeTagger. The use of tokens is a better measure for calculating tagging accuracy.

- Use of optional markers in sentences (e.g. Seit wann has du ihn/sie/es? / *how long have you had him/her/it?*)
- The use of optional male/female endings in one token space (e.g. Wer ist dein(e) Lieblingsschauspieler(in)? / *Who is your favourite actor/actress?*)

In some cases TreeTagger operated correctly for the majority of instances of a listing of words, but then presented one erroneous tag (the full TreeTagger tagset with examples can be found in Appendix M). One example of this was within a listing of the months of the year, TreeTagger correctly tagged 11of the 12 months of the year as 'NN' which is a noun, but then tagged one of them (in the centre of the listing) as 'NE' which is a proper noun. Other errors in tagging included:

- Incorrect tagging of verb forms and sentence particles, like the use of 'bitte' / *please* at the end of a sentence.
- Failing to separate out numbers from anything which surrounded them, for example, €5 was left whole, rather than tagging '€' and '5' separately, and 2:3 was likewise tagged as a whole entity.
- Failing to recognise modal verbs and erroneously using the tag 'VVFIN' (finite verb) instead of the correct tag 'VMFIN' (finite modal verb).
- Incorrect tagging of the word 'Wie' / *how*. The tag 'KOUS' (subordinating conjunction) was applied instead of the correct 'PWAV' (adverbial interrogative pronoun).
- Incorrect tagging of 'bis' / *until* as 'KON' (coordinating conjunction) rather than 'APPR' (preposition in left position (before the Noun Phrase)).

There were no further obvious explanations for other errors made by the tagger. However, as the manually checked file was free from tagging errors, it could be used for the next stage of work in analysing the content. Figure 6.1 above illustrates some of the errors which were made by TreeTagger within the first few sentences of the tagged file (highlighted in red).

While the language contained within the raw linguistic content was quite basic (as beginner's level language for children), it contained many difficult cases for Computational Linguistics (CL)/Natural Language Processing (NLP) systems. It is therefore not surprising, that the tagging system did not perform at its optimum. Taking a few examples from the raw linguistic content and examining the results of Machine Translation (MT) on what seem like 'simple' sentences illustrate this difficulty. For example, the sentence 'W*ie heißt du? / What is your name?*' presents a problem to many machine translation systems. In English we use the copula *is* with two arguments (What [arg1] is [copula] your name [arg2]?). In German the predicate *heißen* has two arguments (Wie [arg1] heißt du[arg2]?). This may cause problems if the correct rule is not in place to translate correctly. It may similarly present problems to other types of CL/NLP tools.

A second example examines the German verb *'machen / to make', which* can present a problem where the correct lexical choice is not made when translating from German to English. This verb is used within certain fixed expressions and has a different meaning within these contexts: *to do (in your free time), to take (a photo)* and *to have (a picnic).*

A third example involves use of the verb '*haben / to have',* which is used differently to its English equivalent. '*Haben / to have'* is used with emotions and conditions of the body, constructions that would usually require the English '*to be*'.

e.g. Ich habe Durst – I am thirsty

Ich habe Angst – I am afraid

This might similarly cause problems for MT systems, if they do not have the fixed expression '*Durst haben / to be thirsty'* in their lexicons. Table 6.1 illustrates the output of two online MT systems for these three examples of difficult cases.

INPUT	**Systran[18] output**	**Reverso[19] output**
Wie heißt du?	How called is you?	How are you called?
What is your name?	Was ist Ihr Name?	Wie heißt du?
Was machst du in deiner Freizeit?	What do you make in your spare time?	What do you do in your leisure?
What do you do in your free time?	Was tun Sie in Ihrer freien Zeit?	Was tun Sie in Ihrer freien Zeit?
Ich habe Hunger	I have hunger	I have hunger
I am hungry	Ich bin hungrig.	Ich habe Hunger

Table 6.1 Output from 2 online MT systems with sample sentences from the raw linguistic content

6.1.3 Post-processing

Following tagging, all of the instances which caused difficulty for the tagger were removed and amended to ensure the data was more readable for further processing during the analysis phase in the next section. Post-processing also ensured that the raw linguistic resources were clearly tagged and formatted for the development and population of the end linguistic content database (see Chapter 7). The analysis stage required that certain parts of speech were automatically selected and counted. The correct identification of all tokens was therefore important. The following amendments were made during post-processing:

- A full stop was added after each of the vocabulary items to mark the end point of the group in each instance, for example, Der Vater. Der Mutter…

[18] Systran available at http://www.systransoft.com/lp/free-online-translation/

[19] Reverso available at http://www.reverso.net/text_translation.aspx?lang=EN

- All ‘pl’ markers on vocabulary items were renamed as ‘plural’ so that they could be better identified as ‘FM’ (foreign language)
- All ‘sing.’ markers on vocabulary items were renamed as ‘singular’ so that they could be better identified as ‘FM’ (foreign language)
- All instances where [] was used to show where text should be inserted were separated out onto their individual lines and the relevant tag was inserted, for example, Er heißt [NE] / *His name is [NE]* where ‘NE’ stands for a proper noun.
- Any missing lemmas from the tagging process were inserted. This meant that all instances of <unknown> instead of the relevant lemma were replaced with the lemma information.
- All optional markers used in sentences were removed and replaced with full sentences in each instance, for example, Ich bin neun/ zehn Jahre alt / *I am nine/ ten years old* was converted to the two sentences, Ich bin neun Jahre alt / *I am nine years old* and Ich bin zehn Jahre alt / *I am ten years old.*

6.2 Analysis and characterisation of raw linguistic content

The background to content analysis was discussed in Section 3.3 (p. 91). The problem space definitions of Section 4.1.4 (p. 119) outlined the relevance of content analysis to this research instance, namely, to examine and analyse the language developed and contained in the raw linguistic content, to compare and contrast that language to other types of language used for didactic purposes and to examine the extent to which the raw linguistic content has features of a controlled language.

The rationale behind this latter goal was to examine whether the language contained within the raw linguistic content exhibited any particular features which made it more or less useful for integrating CL/NLP techniques. CL/NLP techniques have a lot to offer in language learning domains and in Intelligent CALL (ICALL) (e.g. Amaral and Meurers, 2011; Heift and Schulze, 2007; Koller, 2007). Certain subsets of language such as controlled languages or limited domains can work more effectively with CL/NLP technology as the requirement for full-language coverage is reduced.

Along with the raw linguistic content, two other datasets were used for comparative purposes – a corpus comprising newspaper texts in this instance and language used in a textbook for teaching adult learners of German.

Sections 6.2.1 and 6.2.2 examine the three data sources on which the analysis was conducted and the analysis ratios employed (some of these were very briefly examined in the discussion in Section 3.3). The final subsection, Section 6.2.3, documents the results from the analysis.

6.2.1 Comparative Data Sources

Three datasets were used to compare and contrast to one another. These three datasets were (1) the raw linguistic content, (2) the TIGER corpus and (3) Klett's Passwort Deutsch German textbook for adult learners. The TIGER corpus of newspaper texts was chosen as a 'control' dataset; one which could also represent more authentic language than the didactic language of the raw linguistic content and the adult German textbook. The dataset required for comparative purposes only necessitated that it should be a non-didactic dataset.

While a child's German language textbook would have been preferable as the third dataset, none were present in the Irish market which had been developed for primary language learners.[20] Attaining publishers' permission to acquire and use their textbooks as part of this study also proved difficult. Permission could only be sought for an international adult German textbook.

Table 6.2 summarises the data sets used. Each dataset will be considered in turn below alongside the processing which was carried out at various stages to allow the relevant data to be accessed.

Data set		**Description**
1	Raw linguistic content	4,013 tokens long. Contains raw linguistic resources
2	TIGER corpus consecutive samples x10	4,013 tokens long. Each of the 10 randomly and automatically generated consecutive files contains the token followed by the POS information
	TIGER corpus random samples x10	4,013 tokens long. Each of the 10 randomly and automatically generated random files contains a mixture of tokens chosen at random and pieced into a file. Files contain token followed by the POS information
3	Passwort Deutsch adult's German language textbook	4,013 tokens long. This file was created by extracting the first 4,013 tokens from the textbook text supplied by the publisher.

Table 6.2 Description of the 3 data sets used for comparison and analysis

Dataset 1: Raw linguistic content

The raw linguistic content comprised the content under development and described thus far. Its topics and themes were arranged around the content of the modern language curriculum documents.

[20] Only one French language textbook was developed for the primary level cohort in Ireland over the 15 year term of the MLPSI.

The total number of tokens in the dataset after processing was 4,013. Content from the two other datasets – the TIGER corpus and Passwort Deutsch textbook - were selected and trimmed (see below) to match the same running length so that texts were comparable in length. This offset the challenges experienced with the use of alternative analysis matrices to account for text length (as described in Section 3.3).

A Perl script was written to extract the POS information (Appendix N, File Ref. 1) and replace each token within its full sentence on its own line (Appendix N, File Ref. 2). This resulted in a file with one sentence per line. This was the input format for WordSmith tools. The original POS-tagged and tokenised file (one token per line) was used for all other custom analysis processes.

Dataset 2: TIGER corpus

Version 1 of the TIGER treebank (Brants et al, 2004) contained 700,000 tokens (40,000 sentences) compiled from the German newspaper Frankfurter Rundschau. The corpus was semi-automatically tagged with syntactic structures and uses the same markup as TreeTagger. The TIGER corpus represented a corpus-based source of didactic content for the analysis, also representing authentic language (although requiring the mediation outlined in Section 3.2.4 (p. 78) before it could be presented to learners).

The content of the TIGER corpus (see Figure 6.2) contained syntactic structure references which used a numbering scheme to show how elements of a sentence were structurally related (edge, parent and secedge comment). The full TIGER file was trimmed to remove the structural information and dependencies. This was achieved through writing a Perl script to extract these elements and only leave tokens and their respective POS tag (Appendix N, File Ref. 3), to match the format of the raw linguistic content file.

```
#EOT SECEDGETAG
%% word                  tag     morph          edge    parent  secedge comment
#BOS 4 77 1057916823 1 %% @SB2AV@
Der                     ART     --             NK      502
texanische              ADJA    --             NK      502
Milliardär              NN      --             NK      502
Ross                    NE      --             PNC     500
Perot                   NE      --             PNC     500
hat                     VAFIN   --             HD      505
das                     ART     --             NK      503
politische              ADJA    --             NK      503
Establishment           NN      --             NK      503
in                      APPR    --             AC      501
Washington              NE      --             NK      501
aufgeschreckt           VVPP    --             HD      504
.                       $.      --             --      0
#500                    PN      --             NK      502
#501                    PP      --             MNR     503
#502                    NP      --             SB      505
#503                    NP      --             OA      504
#504                    VP      --             OC      505
#505                    S       --             --      0
#EOS 4
```

Figure 6.2 An extract from the TIGER treebank showing all annotation for one sentence (#4)

Ten consecutive or running samples of text were automatically extracted for use through writing a Perl script (Appendix N, File Ref. 4). Each started at a randomly assigned number (marking one point in the corpus) (using the Perl random() function) which was always 4,013 tokens away from the end of the file and extracted 4,013 lines of the corpus from that point onwards. Ten samples were extracted so that any variance across the newspaper genres or topics could be discounted by averaging out the ratio results across the 10 samples. Figure 6.3 provides an extract of one of the consecutive sample files.

```
auf     APPR
ökologischen    ADJA
Anbau   NN
erzielt VVPP
werden  VAINF
.       $.
''      $(
Solche  PIAT
Sätze   NN
weisen  VVFIN
weit    ADJD
über    APPR
betriebswirtschaftliche ADJA
Überlegungen    NN
hinaus  PTKVZ
.       $.
Honoriert       VVPP
wurde   VAFIN
dies    PDS
bisher  ADV
unter   APPR
anderem PIS
mit     APPR
der     ART
Bayerischen     ADJA
Umweltmedaille  NN
,       $,
dem     ART
Titel   NN
Öko-Manager     NN
des     ART
Jahres  NN
,       $,
dem     ART
Umweltpreis     NN
der     ART
Stadt   NN
München NE
und     KON
```

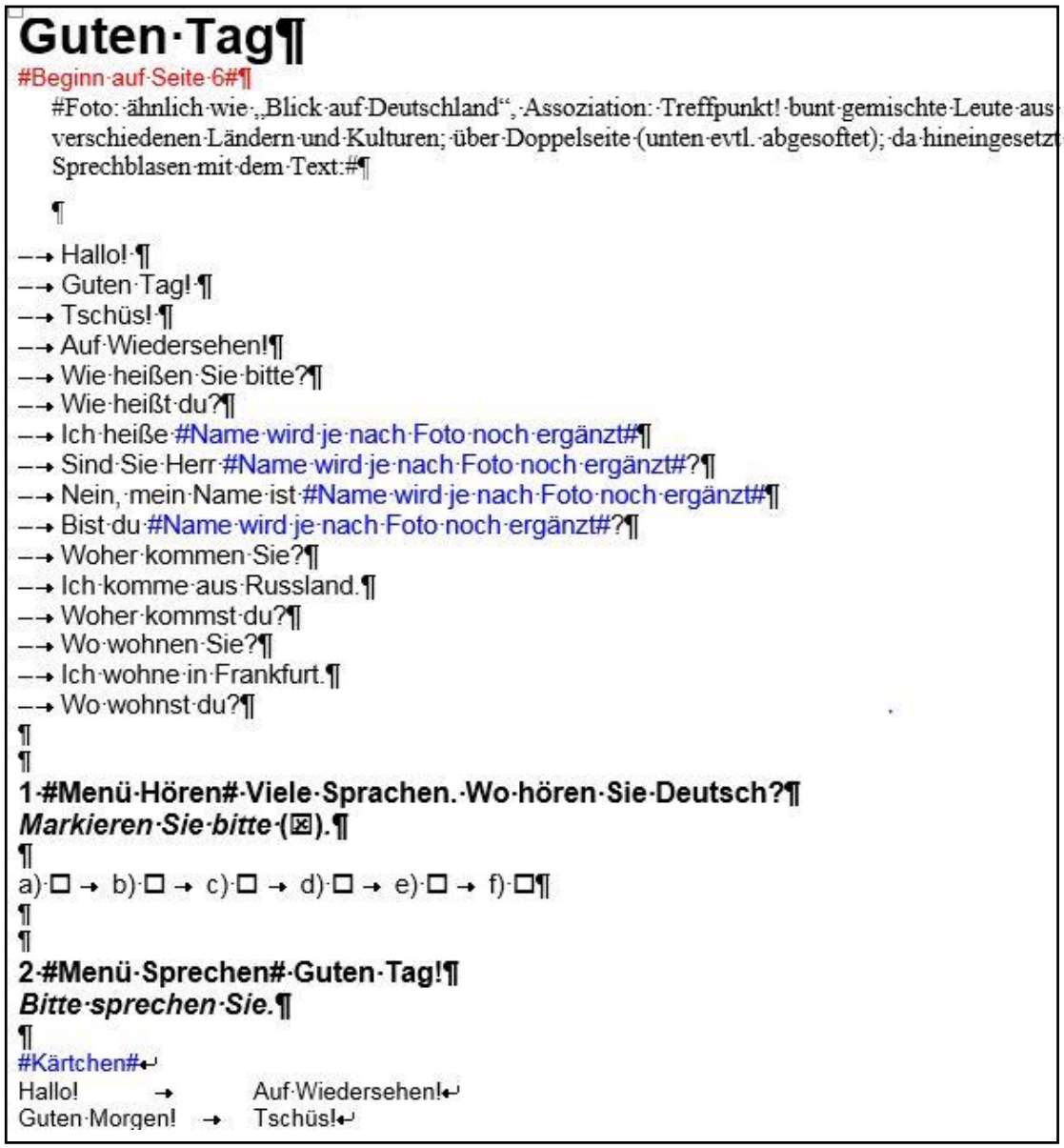
Figure 6.3 Extract from one of the ten consecutive sample files created from the TIGER corpus

Ten random samples of words were also automatically extracted for use using a similar automatic process employing a Perl script (Appendix N, File Ref. 5). The 4,013 tokens were taken from 4,013 random positions and pieced back together to form each of the 10 random sample files. This was completed to see how completely random (albeit newspaper language) text compared to the other datasets, and also to examine how a non-didactic form of random language would measure up against the other datasets. Again, the average of the ratios computed from the 10 random samples was used in each case for comparison. These samples represented the control of the analysis.

Dataset 3: Adult textbook: Passwort Deutsch (Ernst Klett Verlag, 2005)

A shortlist of German adult textbooks was initially created through using the internet to gather information about relevant texts. The main publishing companies (Ernt Klett Verlag, Hueber Verlag and Langenscheidt) were contacted to request (1) permission to use the text of their textbooks for the purposes of this research and (2) access to an electronic file of the first 4,013

lines in any one relevant textbook. Of the three publishers contacted, only Ernst Klett Verlag replied to the correspondence and also granted permission as requested for their Passwort Deutsch[21] textbook. They emailed through an electronic version of the content from the first ten units of the textbook (extract provided in Figure 6.4).



Guten·Tag¶

#Beginn·auf·Seite·6#¶

#Foto:·ähnlich·wie·„Blick·auf·Deutschland“,·Assoziation:·Treffpunkt!·bunt·gemischte·Leute·aus·verschiedenen·Ländern·und·Kulturen;·über·Doppelseite·(unten·evtl.·abgesoftet);·da·hineingesetzt·Sprechblasen·mit·dem·Text:#¶

¶

–→ Hallo!·¶
–→ Guten·Tag!·¶
–→ Tschüs!·¶
–→ Auf·Wiedersehen!¶
–→ Wie·heißen·Sie·bitte?¶
–→ Wie·heißt·du?¶
–→ Ich·heiße·#Name·wird·je·nach·Foto·noch·ergänzt#¶
–→ Sind·Sie·Herr·#Name·wird·je·nach·Foto·noch·ergänzt#?¶
–→ Nein,·mein·Name·ist·#Name·wird·je·nach·Foto·noch·ergänzt#¶
–→ Bist·du·#Name·wird·je·nach·Foto·noch·ergänzt#?¶
–→ Woher·kommen·Sie?¶
–→ Ich·komme·aus·Russland.¶
–→ Woher·kommst·du?¶
–→ Wo·wohnen·Sie?¶
–→ Ich·wohne·in·Frankfurt.¶
–→ Wo·wohnst·du?¶
¶
¶

1·#Menü·Hören#·Viele·Sprachen.·Wo·hören·Sie·Deutsch?¶
Markieren·Sie·bitte·(☒).¶
¶
a)·☐ → b)·☐ → c)·☐ → d)·☐ → e)·☐ → f)·☐¶
¶
¶

2·#Menü·Sprechen#·Guten·Tag!¶
Bitte·sprechen·Sie.¶
¶
#Kärtchen#↵

Hallo!	→	Auf·Wiedersehen!↵
Guten·Morgen!	→	Tschüs!↵

Figure 6.4 Extract from the MS Word file supplied by Ernst Klett Verlag (formatting visible to indicate the level of editing required)

The electronic file which was forwarded was one which had been converted from a designed PDF to MS Word by the publishing company before it was emailed through. This meant that much of the content needed to be heavily reformatted to realign the words of sentences together (sometimes they were stored across separate columns, or in alternative locations on the page (after images had been deleted, for example)). Image references were still embedded in the text

[21] Curtis, H. (2005). *Passwort Deutsch*. Stuttgart: Ernst Klett Sprachen GmbH

and needed to be removed (see blue text in Figure 6.4) and hashtags (#) were used to mark comments in the text. Small Perl scripts were written to assist in some of the file extraction (e.g. to remove all of the hash tags and embedded text – see Appendix N, File Ref. 6). This combined automatic and manual phase of processing took some time to complete as the text required retyping in most instances. It highlights the inaccessibility of the content in its designed form and the importance of maintaining the separation of content from its product to facilitate reuse and editing of content.

Once the text was reformatted into the required format, it was tagged using TreeTagger. The tags assigned were manually checked for accuracy and changes were made where any erroneous tags were encountered. As above, the content was amended automatically via Perl scripts (Appendix N, File Ref. 1 and 2) so that it could be input to WordSmith tools and the full tagged version was used elsewhere with the custom developed Perl scripts.

Table 6.3 provides a full listing of the Perl scripts created for processing and extracting the datasets as described. The full code listing can be found in Appendix N in each case (as identified).

File Ref.	File name	Description
1	Convert_corpus_to_pure_text2.pl	written to extract POS information from a POS-tagged file and replace each token within its full sentence on its own line
2	Convert_words_to_sent_per_line.pl	Written to replace each token within its full sentence on its own line
3	Tiger_preprocessing.pl	Written to extract the token and POS information from the TIGER treebank
4	get_consecutive_sample3.pl	Written to create consecutive random sample files from the edited TIGER corpus (token+POS from 3 above)
5	Get_random_sample.pl	Written to create random sample files (random selections of words together) from the edited TIGER corpus
6	Remove_hash_text_in_textbook.pl	Remove all hashtags and associated comments from the Passwort Deutsch textbook

Table 6.3 Perl scripts created for processing and extracting the datasets as described above

6.2.2 Linguistic analysis ratios employed

Section 3.3 described three domains through which content analysis can be examined

(1) materials development for language teaching and learning,

(2) SLA

(3) Computational Linguistics/Machine Translation.

The first of these referred to cultural analysis and mostly manual examinations of textbook content, and was therefore not relevant to the analysis processes of this chapter. Those used within this research drew on ratios and features outlined from the latter two domains. Some of the SLA ratios were described in Section 3.3 to highlight the variance among ratios to calculate the same metric. The majority of these variations were developed to account for making comparisons between texts of different lengths or adopted by alternative researchers/practitioners (Wolfe-Quintero et al, 1998). As the datasets in this instance were the same length, variation in length was not an issue. However, to avoid doubt within the analyses of the three datasets and to ensure complete coverage, each variation in ratio was used and calculated. The following sections outline the matrices used and calculations made on the datasets according to SLA and CL/MT. Section 6.2.3 then describes the results of the analyses.

SLA Examined: Measures used to examine complexity

In each case below, the ratio, description and process used are outlined. Table 6.4 outlines the Perl scripts which were developed for the analysis of the datasets and the full code listing can be found in Appendix O. In the majority of cases, a Perl script was written to process the language as needed and extract the relevant count or detail for computing ratios (in Excel). For two of the initial ratios (TTR and STTR) data was gathered by running the datasets through the Wordsmith corpus analysis tools (published online by Lexical Analysis Software Ltd. and Oxford University Press). The university of Leipzig Dolch lists were used as the listing of the frequently used words for calculating the sophistication/basicness measures. The top 10, 20 and 30 verbs were also extracted from this file automatically using Perl script ref. 7 (Appendix O).

File Ref.	File name	Description
7	Find_store_adj_noun_verb.pl	Finds all of the adjectives, nouns and verbs in a file and outputs each of the word category listings to their own file (e.g. adjectives tokens, noun tokens etc.)
8	Count_verb_types_from_tokens.pl	Counts the number of types of a particular word class (e.g. verbs) from a listing of relevant tokens.
9	Count_compare_to_freq_lists_output_tokens.pl	Compares the content of a dataset to a defined frequency list input file. Outputs the sophisticated and basic tokens (for use with file 8 to calculate types from tokens)

Table 6.4 Perl processing files to extract and calculate tokens, types and complexity/basicness within datasets

Relevant ratios	Formula	Description	Process
Type/Token Ratio (TTR)	$\frac{Types}{Tokens} \times 100$	Types: Unique words Tokens: Words	WordSmith Tools
	Measures the number of tokens per type (the amount of repetition across a text)		
Standardised TTR (STTR)	Measures the TTR but takes account of the total text length by taking the average TTR of computing the TTR a few times across a certain length of text (e.g. if n=500, the TTR is calculated every 500 words and then the average TTR taken as the STTR.)		WordSmith tools
#Lexical word tokens	Counts the number of noun, adjective and verb tokens		Perl file ref 7
#Lexical word types	Counts the number of noun, adjective and verb types (except auxiliary verbs and modal verbs)		Perl file ref 8
#Verb tokens	Counts the number of verb tokens (by identifying the TreeTagger tags VVFIN, VVIMP, VVINF, VVIZU, VVPP, VAFIN, VAIMP, VAINF, VAPP, VMFIN, VMINF, VMPP)		Perl file ref 7
#Verb types	Counts the number of verb types		Perl file ref 8
#Noun tokens	Counts the number of noun tokens by identifying the TreeTagger tags NN and NE		Perl file ref 7
#Noun types	Counts the number of noun types		Perl file ref 8
#Adj tokens	Counts the number of adjective tokens by identifying the TreeTagger tags ADJA and ADJD		Perl file ref 7
#Adj types	Counts the number of adjective types		Perl file ref 8
Word Variation-1	$\frac{WT}{W}$	WT: Total types W: Total tokens	Excel calc. from Perl extracted data
	Calculates the word variation in the text		
Word Variation-2	$\frac{WT}{\sqrt{2W}}$	WT: Total types W: Total tokens	Excel calc. from Perl extracted data
	Calculates the word variation in the text		
Word Variation-3	$\frac{WT^2}{W}$	WT: Total types W: Total tokens	Excel calc. from Perl extracted data
	Calculates the word variation in the text		
Verb Variation-1	$\frac{VT}{V}$	VT: Verb types V: Verb tokens	Excel calc. from Perl extracted data
	Calculates the verb variation in the text		
Verb Variation-2	$\frac{VT}{LW}$	VT: Verb types LW: Lexical word tokens	Excel calc. from Perl extracted data
	Calculates the verb variation in the text		
Lexical Variation	$\frac{LWT}{LW}$	LWT: Lexical word types LW: Lexical word tokens	Excel calc. from Perl extracted data
	Calculates the lexical (noun, adjective and verb [except auxiliary verbs and modal verbs]) variation in the text		
Noun Variation	$\frac{NT}{LW}$	NT: Noun types LW: Lexical word tokens	Excel calc. from Perl extracted data

Relevant ratios	Formula	Description	Process
	Calculates the noun variation in a text		
Adjective Variation	$\frac{AdjT}{LW}$	AdjT: Adj types LW: Lexical word tokens	Excel calc. from Perl extracted data
	Calculates the adjective variation in a text		
Lexical Density	$\frac{LW}{W}$	LW: Lexical word token W: Total no. tokens	Excel calc. from Perl extracted data
	Calculates the adjective variation in a text		
#Soph. verb types	Calculates the verb types which are not in the most frequently used verbs		Perl file ref 9 Perl file ref 8
#Soph. lexical word tokens	Calculates the lexical word tokens which are not in the most frequently used lexical items		Perl file ref 9
#Soph. word types	Calculates the word types which are not in the most frequently used lexical items		Perl file ref 9 Perl file ref 8
#Basic word types	Calculates the word types which are in the most frequently used lexical items		Perl file ref 9 Perl file ref 8
Verb Sophistication	$\frac{SVT}{V}$	SVT: Soph. verb types V: Verb tokens	Perl file ref 9 Perl file ref 8 Excel calc. from Perl extracted data
	Calculates the sophistication of verbs (those which are not in the most frequently used listing)		
Verb Sophistication 2	$\frac{SVT^2}{V}$	SVT: Soph. verb types V: Verb tokens	Perl file ref 9 Perl file ref 8 Excel calc. from Perl extracted data
	Calculates the sophistication of verbs (those which are not in the most frequently used listing)		
Verb Sophistication 3	$\frac{SVT}{\sqrt{2V}}$	SVT: Soph. verb types V: Verb tokens	Perl file ref 9 Perl file ref 8 Excel calc. from Perl extracted data
	Calculates the sophistication of verbs (those which are not in the most frequently used listing)		
Lexical Sophistication 1	$\frac{SLW}{LW}$	SLW: Soph. lexical word tokens LW: Lexical word tokens	Perl file ref 9 Excel calc. from Perl extracted data
	Calculates the sophistication of lexical items (nouns, adjectives and verbs (excluding auxiliary verbs and modal verbs) (those which are not in the most frequently used listing)		
Lexical Sophistication 2	$\frac{SWT}{WT}$	SWT: Soph. word types WT: Word types	Perl file ref 9 Perl file ref 8 Excel calc. from Perl extracted data
	Calculates the sophistication of lexical items (nouns, adjectives and verbs (excluding auxiliary verbs and modal verbs) (those which are not in the most frequently used listing)		
Lexical Basicness	$\frac{BWT}{WT}$	BWT: Basic word types WT: Word types	Perl file ref 9 Perl file ref 8 Excel calc. from Perl extracted data
	Calculates the basicness of lexical items (nouns, adjectives and verbs (excluding auxiliary verbs and		

Relevant ratios	Formula	Description	Process
	modal verbs) (those which are in the most frequently used listing)		

Table 6.5 Analysis ratios relevant to this research

The ratios listed in Table 6.5 were all used to examine the complexity of the three datasets. As mentioned, the use of more than one ratio in the calculation of a measure (e.g. word variation) was to account for variation in text length. As all of the sample datasets were the same length, each of the alternative ratios were calculated for completeness. The initial listings, such as #verb tokens and #noun tokens were outlined and calculated as their values formed part of the more complex ratios which were to be calculated.

CL/NLP and controlled languages

As mentioned in Section 3.3 (p. 91), there is much scope for the automatic analysis of language learning content. This can be carried out using CL/NLP systems. Overcoming the challenges surrounding coverage, robustness and accuracy allow CL/NLP systems to be used more successfully. Table 6.6 outlines the Perl scripts which were created to calculate elements relating to CL/NLP which are described below. The full code listing can be found in Appendix O.

File Ref.	File name	Description
10	Count_punc_per_sent.pl	Counts the number of punctuation marks in a file to be divided by the number of sentences in the file (indicated by sentential-final punctuation such as . , ! ? : ;).
11	Count_store_conj_per_file.pl	Counts and stores the number of conjunctions per file to be divided by the number of sentences in the file.
12	Count_store_rel_pron_per_file.pl	Counts and stores the number of relative pronouns per file to be divided by the number of sentences in the file.
13	Reading_into_array.pl	Computes the average word length in a file (to nearest integer)

Table 6.6 Perl processing files created to extract and calculate elements indicating the presence of controlled languages

The features described in Section 3.3.3 in respect controlled languages and controlled language checkers indicated the types of rules which are conformed to within controlled languages where they are restricted by grammar and lexical rules. These controlled language rules simplify a text. Hence, the level of complexity in each sentence in a corpus or set of test data can be examined by analysing some of these controlled language features. Those features of controlled languages which were used to characterise the datasets have been compiled in Table 6.7. They help to indicate the level of complexity or basicness of the content of the datasets. For example, if a

sentence contains more than one punctuation mark (the final punctuation mark to indicate the end of the sentence), then it will be more complex than a sentence which only contains one punctuation mark as it will likely include a conjunction and additional phrases, making it more complex. In the same way, higher numbers of conjunctions and relative pronouns in a sentence indicate greater complexity than lower numbers of them. The measures listed in Table 6.7 were examined in the case of the first and second phases of testing as outlined in Section 6.2.3 below.

Relevant controlled language indicators	**Process**
Number of tokens	WordSmith tools
Number of types	WordSmith tools
Type/token ratio (TTR)	WordSmith tools
Standardised TTR (STTR) (TTR which accounts for text length)	WordSmith tools
Average word length	Perl file ref 13
Average sentence length	Perl file ref 10 Excel calc. from Perl and WordSmith extracted data
Number of punctuation marks	Perl file ref 10
Number of co-ordinating & subordinating conjunctions	Perl file ref 11
Number of relative pronouns	Perl file ref 12
Frequency comparison with 100 most frequent words in German	Perl file ref 9

Table 6.7 Features of controlled languages which were examined and calculated

6.2.3 Analysis results and comparison of data sources

The SLA analysis ratios were used for comparing the complexity of the datasets to one another and the measures of general features of controlled languages were tested across the three data sources during two phases of testing.

The **first phase of testing** (see complete results in Appendix Q) using the SLA and CL matrices tested the outcomes on all of the datasets in the format outlined in Section 6.2.1. As the testing progressed, the results started to reveal that the non-complete sentences of vocabulary items in the data sets were resulting in a skewing of the complexity measures as the lexical items were heavier than would usually be the case in running sentences of language (as lists of vocabulary items comprising nouns, adjectives or verbs were adding extra bulk to these ratio calculations). As a result, a second phase of testing was carried out on revised datasets where these bulky items were removed.

Phase two of testing (see complete results in Appendix R) using the SLA and CL matrices was conducted on an abbreviated set of data (totalling 2,496 tokens) where the raw linguistic content and the German language textbook were processed to remove these bulky lexical items which

did not equate to full sentences (and result in more accurate complexity measure calculations). The following types of text were removed from the raw linguistic content:

- English headers and section names
- Lists of adjectives (e.g. weather, traits, colours)
- Lists of nouns (e.g. school subjects and types of music)
- Lists of verbs (e.g. relating to learning at school)
- ADJA NN (adjective noun) combinations which did not form full sentences
- Lists of numbers
- Dates (e.g. FM ADJA – foreign language adjective – 2nd zweite).

Similarly, incomplete sentences from the German language textbook were also deleted as follows:

- ADJA NN (adjective noun) or ART NN (article noun) combinations which did not form full sentences
- NE (proper nouns) (e.g. country names)
- NN (nouns) (e.g. letters of the alphabet)
- ADJD (adjectives) (e.g. richtig / *true*, falsch / *false*, singular, plural)
- Lists of random words which are mixed up for the learner to form correct sentences from
- Lists of verbs (e.g. what you do on the train)
- Lists of numbers
- Lists of dates.

Once these items were removed from the raw linguistic content file and the German language textbook file, the TIGER corpus files and the German language textbook sample files were trimmed to match the same length as the amended raw linguistic content file (2,496 tokens).

The full results are detailed in Appendix Q and R. The main and most important ratios from both phases of testing are summarised below.

1. PHASE ONE TESTING

The analyses carried out on the test datasets in examining both the level of complexity of the content and the conformity to controlled language rules revealed similarities and differences across the characteristics of the datasets. The controlled language rules are examined below first, as these provide a basis for the complexity measures from SLA which follow.

	Lexical analysis					Syntactic analysis			
TEST DATA	**#Tokens**	**#Types**	**TTR**	**STTR**	**Average word length (chars.)**	**Average sent. Length (words)**	**#Pun. Marks per sent. (incl. final)**	**#Co-ord. conjs.**	**#Subordinating Conjs.**
Raw linguistic content	2,887	909	32.23	36.25	4.84	3.18 (skewed)	1.15	3	1 (eg. in 4 sens.)
TIGER consecutive samples	3,610.7	1,573.5	44.46	54.98	6.12	17.86	2.03	98.2	50.4
TIGER random samples	3,601.1	1,738	49.07	59.68	6.17	---	---	86.5	47.8
Passwort Deutsch Textbook	3,239	540	18.11	26.10	4.58	5.14 (skewed)	1.32	92	0

Table 6.8 Testing phase one: Examination of lexical and syntactic conformity to generic controlled language rules

Testing for controlled language features

When the raw linguistic content was compared to both consecutive and random word samples from the TIGER corpus, the results were a lot lower in all instances indicating that the language in the raw linguistic content was more controlled than the language in both the consecutive and random samples from the TIGER corpus. Lexical analysis (see Table 6.8) showed that the raw linguistic content had lower TTR and STTR values (TTR: 32; STTR: 36) than the TIGER samples (TTR: 44/49 and STTR: 54/59), indicating lower lexical complexity. Similarly, the average word length of the raw linguistic content was 4.8 which was lower than the 6 character per word average of the TIGER samples. Shorter sentences and reduced lexical complexity are features of controlled languages.

Syntactic indicators of controlled languages also indicated lower levels of syntactic complexity in the raw linguistic content than in the TIGER samples. The average sentence length for the raw linguistic content was very low, at 3 words per sentence, in comparison to the 18 word sentences of the TIGER samples (the random samples would not have any naturally occurring syntactic form as the words were randomly pulled together). The raw linguistic content sentence length was somewhat skewed as it contained many listings of "article noun" pairs, which offset the true sentence lengths. This skewing was addressed through the content trimming of Phase 2 testing (see below). The raw linguistic content had half the number of punctuation marks that the TIGER samples did, and only about 3% of the number of conjunctions present in the TIGER samples. Similarly, the raw linguistic content did not contain any relative pronouns in

comparison to the 27 and 29 found across the TIGER samples (see Table 6.9). All of these features indicate a lower level of complexity in the raw linguistic content than in the TIGER samples, and a greater level of controlled language evidence. This was an expected outcome given the development methods and level of language defined by the local context. Nevertheless, examining these controlled language features asserted the controlled language characteristics on the raw linguistic content.

Syntactic analysis: Number of relative pronouns			
TEST DATA	**#Substituting relative pronouns (PRELS)**	**#Attributing relative pronouns (PRELAT)**	**TOTAL # Relative pronouns**
Raw linguistic content	0	0	0
TIGER consecutive samples	25.2	1.4	26.6
TIGER random samples	27.3	2.1	29.4
Passwort Deutsch Textbook	0	0	0

Table 6.9 Testing phase one: Examination of syntactic conformity to generic controlled language rules

When the language in the raw linguistic content was also compared with controlled language rules, it was shown to conform to some lexical and syntactic rules from various controlled language rule-systems (where the TIGER corpus language did not). While the raw linguistic content development was not governed by any rules or restrictions outside the modern language curriculum documents (as controlled languages usually are), it does represent a controlled language in its limitations. It represents a simplified subset of language. The language in the raw linguistic content also exhibited properties of both human-oriented (HO) and machine-oriented (MO) Controlled Language.

- HO: simple, non-complex language – learner curriculum for children, needs to be easy and comprehensible as well as being relevant to the language they know in their L1
- MO: the language is easier for a machine to process/parse – less complexity, sentential markers, extra formatting (punctuation, spacing, section markers...)

When the raw linguistic content was compared to the Passwort Deutsch (PWD) textbook, the textbook exhibited lexical features which were more controlled than the raw linguistic content, and syntactic features which were marginally more complex. A more detailed examination of the language in the textbook revealed that only three topics were covered in the 4,013 lines of

the textbook examined in contrast to the 19 covered in the raw linguistic content over the same number of tokens. This was also indicated in the lower number of 'types' in the PWD textbook than for the raw linguistic content. The raw linguistic content represented a '*concentrated*' set of didactic language which is to be applied and used in a more '*diluted*' form. When the raw linguistic content is accessed for use in a learning situation, it will be spread across many lessons and activities, to ensure recycling and repetition of language to be learned. The German language textbook is already presented in this '*diluted*' format and is a resource to be applied directly in teaching and learning.

However, both didactic sources (the raw linguistic content and the PWD textbook) were more controlled than the corpus-based text containing more standard (newspaper) text. When both the raw linguistic content and the PWD textbook were compared to the two TIGER samples, they both prove to be more controlled according to the analysis carried out and a lot less complex.

Testing using ratios from SLA

The complexity measures from SLA showed similar results to those examined for controlled language features above. The word variation, verb variation, lexical variation, noun variation and adjective variation was lower for the raw linguistic content than for the TIGER samples, showing a lesser vocabulary in use in each case. As before, the more diluted textbook showed lower scores for each of the ratios than the raw linguistic content did, meaning the vocabulary in use was less complex than that contained in the raw linguistic content. Both didactic texts showed lower scores than the TIGER corpus samples. Table 6.10 below outlines the scores obtained across the datasets.

MEASURES	**Raw linguistic content**	**Consecutive Sample Average**	**Random Sample Average**	**Passwort Deutsch Textbook**
Word Variation-1 **WT/W** Tot. types/total tokens	0.31	0.44	0.48	0.17
Word Variation-2 **WT/√2W**	11.96	37.03	40.95	6.71
Word Variation-3 **WT(WT)/W**	286.21	685.71	838.81	90.03
Verb Variation-1 **VT/V** Verb types/verb tokens	0.27	0.63	0.62	0.16
Verb Variation-2 **VT/LW** Verb types/lexical word tokens	0.10	0.17	0.17	0.05
Lexical Variation **LWT/LW** Lexical word types/lexical word tokens	0.59	0.73	0.81	0.28
Noun Variation **NT/LW** Noun types/lexical word tokens	0.40	0.43	0.51	0.21
Adjective Variation **AdjT/LW** Adj types/lexical word tokens	0.11	0.15	0.15	0.03
Lexical Density **LW/W** Lexical word tokens/total no. tokens	0.39	0.49	0.50	0.44

Table 6.10 Testing phase one: Examination of complexity (i)

The word variation across each of the three ratios illustrates how the raw linguistic content had lower word variation results when compared to the TIGER samples. Taking the Word-Variation-1 ratio, the raw linguistic content had a result of 0.31 in comparison to the greater results of 0.44 and 0.48 of the TIGER datasets. This meant that the raw linguistic content contained simpler language which repeated more often than the more varied TIGER samples. The PWD results were lower than the raw linguistic resources at 0.17. As discussed, the more diluted form of language contained more repetition and less variation than the more concentrated raw linguistic resources.

The verb variation provided the greatest difference in results between the raw linguistic content and the TIGER samples. The raw linguistic content contained half as much verb variation than the TIGER samples (ca. 3:6 and 1:2). The PWD textbook had the lowest verb variation across the datasets. The lexical variation (6:7) and lexical density (4:5) were all lower for the raw

linguistic content than the TIGER samples. The noun variation and adjective variation results were closer, which was accounted for through the vocabulary-heavy content of the raw linguistic resources. In all cases, the diluted language form of the PWD textbook had the lowest levels of variation. The results obtained were broadly in line with what was expected, with the raw linguistic content having a lower verb variation than the TIGER sample set, and the less condensed PWD textbook having the lowest level of verb variation across its contents.

		Verb Sophistication 1	**Verb Sophistication 2**	**Verb Sophistication 3**
Raw linguistic content	20 freq	0.25	26.44	3.64
	100 freq	0.21	19.97	3.16
	200 freq	0.2	16.67	2.89
TIGER con. samples	20 freq	0.6	170.6	12.05
	100 freq	0.5	123.28	10.25
	200 freq	0.45	99.62	9.21
TIGER ran. samples	20 freq	0.6	168.02	9.16
	100 freq	0.5	119.75	9.49
	200 freq	0.44	93.60	6.84
Passwort Deutsch Textbook	20 freq	0.16	12.04	2.45
	100 freq	0.13	7.97	2.0
	200 freq	0.12	6.48	1.8

Table 6.11 Testing phase one: Examination of complexity (ii): Verb sophistication (red text indicates the read order for correlating data)

Complexity measures were calculated by comparing the data sets to the university of Leipzig Dolch lists. These Dolch lists contain the most frequently used words in German. Comparing their content to the content of the datasets provides an indication of the level of complexity (where there are few words in common) or basicness (where there are many words in common) of the contents of the datasets. Dolch listings were trimmed to 20-, 100- and 200-word lengths to ascertain the levels of complexity across the datasets. Table 6.11 shows that the raw linguistic content represented a lower verb sophistication level than the TIGER samples, as the raw linguistic content had much crossover with the listing of the most frequently used (and therefore, unsophisticated) words. The raw linguistic content was half as sophisticated as the TIGER samples (ca. 1:2) for the verb sophistication 1 ratio, 6 times less sophisticated with the second ratio (ca. 1:6) and 3 times less sophisticated with the third ratio (ca. 1:3). The PWD textbook again proved the least sophisticated of all of the datasets.

The differences between the raw linguistic content and the TIGER samples were apparent when the lexical sophistication ratios were calculated (see Table 6.12 below). In each case, the raw

linguistic content result was higher (more sophisticated) than the TIGER samples. The Dolch list sample was changed to sizes of 1000 and 2000 for these tests, as lexical choice is more varied than verbal (as above). The raw linguistic content contained many vocabulary and adjectival listings in its content, resulting in a skewed lexical result which will be addressed for phase 2 of testing. For lexical sophistication ratio 1, the raw linguistic content was most sophisticated and the PWD book the least. Conversely, for lexical sophistication ratio 2, the figures are reversed but the result pertain to the same ranking of sophistication for each of the datasets as for the first ratio. The TIGER samples were most sophisticated (0.69/0,76), followed by the raw linguistic content (0.85) and finally the PWD textbook was least sophisticated (0.91). The lexical basicness ratio places the raw linguistic content as being the least basic of the datasets (0.21), the TIGER samples being second-most basic (0.27/0.26) and the PWD textbook as being the most basic.

Aside from the make-up of the raw linguistic resources being lexically heavy, the unexpected results could also pertain to the listing of basic words used, the Dolch listing. McCarten and McCarthy (2010) cautioned against the use of frequency listings for language learning as they often do not reflect the usual usage of language. Dolch listings tend to have few content words in the upper end of their listings. For example, the top ten words from the Leipzig listing are der/*the* (masc. nominative case), die/*the* (fem.), und/*and*, in/*in*, den/*the* (masc. accusative case), von/*from*, zu/*to* or separable prefix, das/*the* (neut.), mit/*with* or separable prefix and sich/*himself/herself/itself/etc*. reflexive pronoun. A similar caution could be evidenced here as everyday language (defined in Dolch listings) would not have much overlap with the content of a beginner's language learning course. To this end, there would be little overlap between the two data files, which would result in ratios indicating that the content was very sophisticated.

		Lexical Sophistication 1 **SLW/LW** Soph lex word tokens/lex word tokens	Lexical Sophistication 2 **SWT/WT** Soph word types/word types	Lexical Basicness **BWT/WT** Basic word types/word types
Raw linguistic content	1000 freq	0.82	0.85	0.21
	2000 freq	0.78	0.79	0.28
TIGER con. samples	1000 freq	0.76	0.69	0.27
	2000 freq	0.71	0.27	0.36
TIGER ran. samples	1000 freq	0.80	0.76	0.26
	2000 freq	0.75	0.68	0.34
Passwort Deutsch Textbook	1000 freq	0.25	0.91	0.30
	2000 freq	0.23	0.81	0.41

Table 6.12 Testing phase one: Examination of complexity (ii): Lexical sophistication (red text indicates the read order for correlating data)

2. PHASE TWO TESTING

For this phase of testing, the raw linguistic content and the Passwort Deutsch textbook were all trimmed to remove the listings of vocabulary items which did not form complete sentences. This was done to examine whether the lexical and verb complexity/sophistication measures would be reduced.

With the removal of pure vocabulary items (article-noun combinations, listings of adjectives etc.), the average sentence length of the raw linguistic content was increased to 4.05 words in comparison to the 17.69 word per sentence average of the TIGER samples. The raw linguistic content figure is slightly higher than the textbook content (see Table 6.13). The type:token ratios were also reduced, as the lexical makeup of the raw linguistic resources and PWD textbook were more balanced with full sentences (fewer unique tokens to make up additional types). The numbers of punctuation marks, coordinating and subordinating conjunctions were reduced across the datasets with their reduced content, but retained the same ranking as phase one testing results with the raw linguistic content conforming to controlled language rules (containing fewer of each of these elements) than the TIGER samples, and the diluted form of the PWD textbook containing the least of these elements.

Lexical analysis						Syntactic analysis			
TEST DATA	**#Tokens**	**#Types**	**TTR**	**STTR**	**Average word length (chars.)**	**Average sent. Length (words)**	**#pun. Marks per sent. (incl. final)**	**#Coord. conjs.**	**#Subord. Conjs.**
Curriculum corpus	1,880	467	25.07	27.40	4.51	4.05	1.04	2	1 (eg. In 4 sens)
TIGER consecutive samples	2,252.7	1,064.1	48.19	54.68	6.06	17.69	2	59.1	28.6
TIGER random samples	2,240.4	1,140.8	51.73	58.31	6.19	---	---	55.8	29.3
Passwort Deutsch Textbook	1,839	344	18.88	20.60	4.75	3.93	1.14	78	0

Table 6.13 Testing phase two: Examination of lexical and syntactic conformity to generic controlled language rules

Removal of vocabulary items also resulted in the shortening of the sample size across all of the datasets. As a result, the number of relative pronouns across the TIGER samples was reduced, and naturally remained at 0 for the didactic datasets, as in testing phase one (see Table 6.14 below). These results again highlight the conformance of the raw linguistic content and PWD textbook to controlled language rules (resulting in reduced sentential complexity).

Syntactic analysis: Number of relative pronouns			
TEST DATA	**#Substituting relative pronouns (PRELS)**	**#Attributing relative pronouns (PRELAT)**	**TOTAL # Relative pronouns**
Curriculum corpus	0	0	0
TIGER consecutive samples	15.1	0.9	16
TIGER random samples	17	1.2	18.2
Passwort Deutsch Textbook	0	0	0

Table 6.14 Testing phase two: Examination of syntactic conformity to generic controlled language rules

After reducing the datasets, the word, verb, lexical, noun and adjective variation ratios have all increased but still hold the ranking across the datasets as described above. That is, the raw

linguistic content shows less variation than the TIGER samples and more variation than the PWD textbook.

MEASURES	**Raw linguistic content**	**Consecutive Sample Average**	**Random Sample Average**	**Passwort Deutsch Textbook**
Word Variation-1 **WT/W** Tot. types/total tokens	0.40	0.47	0.51	0.19
Word Variation-2 **WT/√2W**	7.62	15.83	17.00	5.67
Word Variation-3 **WT(WT)/W**	116.00	2252.7	580.89	64.35
Verb Variation-1 **VT/V** Verb types/verb tokens	0.25	0.68	0.65	0.17
Verb Variation-2 **VT/LW** Verb types/lexical word tokens	0.16	0.19	0.18	0.07
Lexical Variation **LWT/LW** Lexical word types/lexical word tokens	0.51	0.79	0.82	0.30
Noun Variation **NT/LW** Noun types/lexical word tokens	0.27	0.47	0.51	0.22
Adjective Variation **AdjT/LW** Adj types/lexical word tokens	0.10	0.17	0.16	0.03
Lexical Density **LW/W** Lexical word tokens/total no. tokens	0.36	0.47	0.49	0.49

Table 6.15 Testing phase two: Examination of complexity (i)

An examination of the ratio results from the reduced datasets reveals similar ratios to before. The word variation using the first ratio of raw linguistic content to TIGER samples is ca. 4:5. The PWD textbook result is half that of the raw linguistic content, at 0.19. Verb variation 1 reveals ratio results showing that the TIGER samples have three times more variety than the raw linguistic resources in their lexical makeup. The PWD textbook has the lowest variation, at ca. ¼ of the verb variation shown in the TIGER samples. The ratio of lexical variation between the raw linguistic resources, the TIGER samples and the PWD textbook was 5:8:3, showing the PWD textbook containing the least amount of variation in verbs, and the raw linguistic content showing the second least amount. Noun variation showed a ratio between the samples of 3:5:2 and adjective variation of 1:2:0.3. The lexical density of the datasets showed the raw linguistic

content as being the least dense (0.36), followed by the PWD textbook (0.49) and the TIGER samples (0.47/0.49) were the most dense (complex).

These results all indicate that the two didactic sources were substantially less varied in their content than the TIGER samples.

Table 6.16 shows that the raw linguistic content is less sophisticated following trimming of the datasets. In the majority of cases, the raw linguistic content is again verbally less sophisticated than the TIGER samples, and more sophisticated than the textbook.

		Verb Sophistication **SVT/V** Tot. soph verb types/total verb tokens	Verb Sophistication 2 **SVT^2/V** Tot. soph verb types squared/total verb tokens	Verb Sophistication 3 **$SVT/\sqrt{2V}$** Tot. soph verb types/sq root 2*total verb tokens
Raw linguistic content	20 freq	0.23	22.76	3.37
	100 freq	0.20	16.72	2.89
	200 freq	0.18	14.05	2.65
TIGER con. samples	20 freq	0.62	117.70	7.67
	100 freq	0.52	81.42	6.38
	200 freq	0.46	64.67	5.69
TIGER ran. samples	20 freq	0.59	104.38	7.22
	100 freq	0.48	69.93	5.91
	200 freq	0.42	53.16	5.16
Password Deutsch Textbook	20 freq	0.17	9.87	2.22
	100 freq	0.14	6.48	1.80
	200 freq	0.12	5.17	1.60

Table 6.16 Testing phase one: Examination of complexity: Verb sophistication (red text indicates the read order for correlating data)

The figures for each of the ratio results have been reduced, as each dataset is shorter, it is also less complex (there is a smaller pool to work from). The verb sophistication of the raw linguistic content is nearly a third that of the TIGER samples at 0.23 with the top 20 most frequently verbs in comparison to 0.62/0.59. The PWD textbook has a result of 0.17, which is closer to the raw linguistic content result of 0.23 than in phase one testing. The results of the verb sophistication measures 2 and 3 follow the same trend as outlined for measure 1.

		Lexical Sophistication 1 **SLW/LW** Soph lex word tokens/lex word tokens	Lexical Sophistication 2 **SWT/WT** Soph word types/word types	Lexical Basicness **BWT/WT** Basic word types/word types
Raw linguistic content	1000 freq	0.82	0.72	0.31
	2000 freq	0.77	0.65	0.39
TIGER con. samples	1000 freq	0.82	0.74	0.30
	2000 freq	0.77	0.67	0.38
TIGER ran. samples	1000 freq	0.84	0.73	0.29
	2000 freq	0.80	0.67	0.36
Password Deutsch Textbook	1000 freq	0.77	0.73	0.37
	2000 freq	0.63	0.63	0.49

Table 6.17 Testing phase one: Examination of complexity: Lexical sophistication (red text indicates the read order for correlating data)

Phase one testing indicated that the results of the lexical sophistication ratios placed the raw linguistic content as lexically more sophisticated that any of the other datasets. Following the removal of partial sentences, the raw linguistic content with a result of 0.82 now comes nearer par with the TIGER samples at 0.82/0.84 and only slightly higher than the PWD textbook at 0.77 with the 1000 most frequently used words. Using the lexical sophistication ratio 2, the raw linguistic content is slightly lower (0.01-0.02 lower) than the other two datasets, showing that it is lexically less sophisticated. The final ratio examining lexical basicness shows the results as the inverse of the previous ratios. The raw linguistic content at 0.31is now more basic than the TIGER samples at 0.30/0.29. The PWD textbook is the most lexically basic at 0.37.

6.3 Implications of the analyses results

The discussion now turns to the fourth research question which queries whether the raw linguistic content, bounded by a curriculum, exhibited features of a controlled language. The answer can only be substantiated through the results of the SLA and CL analyses of this chapter.

The language in the raw linguistic content **is more controlled and less complex than the language in the TIGER corpus**. The raw linguistic content had shorter words, shorter sentences, fewer punctuation marks, substantially fewer coordinating conjunctions and substantially fewer subordinating conjunctions than the TIGER samples. Similarly, it did not contain any relative pronouns. All of these features are features of controlled languages, which aim to simplify texts according to grammar and lexical rules.

The raw linguistic resources exhibited lower word, verb, lexical, noun and adjective variation than the TIGER samples. The raw linguistic resources were therefore a simpler form of language. Similarly, the lexical density of the raw linguistic content was the lowest of all of the datasets.

However, while the language in the raw linguistic content is verbally less sophisticated than the language in the TIGER corpus, it is only lexically as sophisticated as the TIGER corpus, or marginally lower, depending on the ratio employed. These latter results were improved in the phase 2 testing when non-complete sentences were removed from the didactic datasets and the TIGER samples trimmed to match the new shorter sample length. As discussed, the use of Dolch lists as a source of 'simple' language for comparison to the datasets may have led to misleading results. Further analyses could use an alternative source of 'basic' or 'sophisticated' content for comparison.

The language in the Passwort Deutsch textbook is more controlled than the raw linguistic content and less complex. This is accounted for by the fact that the language analysed in the raw linguistic content represents a 'concentrated' form of didactic language which covers 12 topics to the 3 covered in the more 'diluted' Passwort Deutsch textbook language running to the same length. A more balanced and comparative analysis could examine the outputs of the linguistic content database employing the raw linguistic resources to the PWD textbook.

Both the raw linguistic content and the PWD textbook represented didactic content for beginners. They both proved more controlled and less complex than the TIGER consecutive and random samples.

What are the characteristics of the language within the raw linguistic content?

The language within the raw linguistic content is simple, with short words and short sentences. It is characterised by usually having sentences containing one phrase (with minimal use of conjunctions and no use of relative pronouns). It conforms to controlled language rules. There are low levels of word, verb, lexical, noun and adjective variation when compared to non-didactic language, meaning that the same words are repeated across the content.

Controlled and Sub-Languages in the Primary School Modern Language Context

The primary school environment itself can be argued to exhibit properties commonly associated with controlled language, limited domains or sublanguages. Compared to an adult speaker, primary school children exhibit a limited production performance in their first language (L1) – they have near full competence but their active performance is restricted in terms of vocabulary

and certain constructions (Brown, 1994). In addition, their world knowledge is less than an adults', so that available context tends to be more restricted. 'Limited' L1 production performance constrains L2 performance – students will not tend to be able to produce complex linguistic constructions in the target language when they do not use them in their native language. Limited L1 also reduces invalid *transfer* of complex structures from the L1 to the L2 (constructions which are valid in the learners' L1 but erroneous in the L2).

The language content in the raw linguistic content has been shown to have features of a controlled language. It is confined by its association to the bounds of the modern languages curriculum documents and its 19 topics (e.g. Family, Parts of the Body, and Introducing yourself). The language listing is finite (as delimited in Appendix D). It is also confined by being tailored to children and their reduced linguistic repertoire, as above. These topics constitute limited domains. The language within each of these topics has been detailed resulting in sublanguages. Sentences will be short exhibiting a limited amount of syntactic complexity and variation. Vocabulary is relatively simple.

What are the implications of the raw linguistic content exhibiting features of a controlled language?

Section 3.3 highlighted how CL/NLP tools suffer from challenges of coverage, robustness and accuracy. In instances where one or more of these challenges can be overcome, the result is a successful CL/NLP tool. The examples provided showed how the Météo weather translation system has a confined language which can be fully accounted for in translation (reduced coverage). Similarly, the successful ICALL systems used in university settings which were described by Amaral and Meurers (2011) all reduce the coverage required by narrowing and ensuring the learner input can be accounted for.

Controlled languages are already used in MT to improve the quality of the translation process and to reduce the pre- and post-editing requirements. If the language within the raw linguistic content is similarly controlled, then it too has a reduced coverage and could result in the successful integration and use of CL/NLP tools.

6.4 Summary and conclusion

The raw linguistic content was processed and had POS-tags added automatically by using the TreeTagger CL tool. These POS tags were used to identify certain parts of speech (e.g. nouns, verbs, pronouns) which were used to count the relevant words throughout the analysis tasks.

Analysis of the raw linguistic content revealed that it does exhibit characteristics of controlled languages in its simplified structure and lexical make-up. The Passwort Deutsch textbook exhibited similar characteristics to a greater degree, which were ascribed to its more diluted form of covering 3 topics across the same word count as the raw linguistic content covered a concentrated 19 topics. Controlled languages present greater capacity for CL tools to work more effectively by eliminating the difficulty with language coverage.

The next chapter illustrates how all of the relevant data gathered during time in the research space was compiled and added to the raw linguistic content to form a linguistic content database. It also shows how the content can be put to use by the practitioner in preparing resources for their learners.

Chapter 7 Developing, populating and using the linguistic content database

This chapter describes the final stages of the research design in developing, populating (Phase III) and using the linguistic content database (Phase IV). Section 7.1 examines the technological decisions which were made to choose a suitable mark-up language to use in the database creation. The database specification of Chapter 4 is also revisited in this section and mapped to a database solution. It describes how all of the relevant raw linguistic content and Localised Setting (LS) notes are consolidated and translated into the database structure.

Section 7.1 examines the creation of the linguistic content database (Phase III) and section 7.2 illustrates how it can be used in practice by the end user (Phase IV), such as the practitioner or publishing house. It asserts the reusability of the linguistic content database by showing how the content can be selected and applied into practical teaching and learning resources.

7.1 Converting raw linguistic content to a linguistic content database

This section describes the need to convert the raw linguistic content into a more usable and computer-accessible format to allow its contents to be selected, manipulated and output into alternative applications. The processes outlined for Phase III are outlined in the following subsections and include selecting the relevant database/markup format, consolidating the data to be transferred to the database and specifying the database content.

7.1.1 Researching and selecting the relevant format

There were four main requirements mapped out for the end database system, three of which (2-4) carry through as the core research requirements. Chapters 2-4 outlined that the linguistic content database should:

1. be accessible to the end user for the quick and easy development of resources
2. account for the localisation of resources through its structure and content
3. allow for reusability and transferability of content
4. ensure the separation of content from its product.

These requirements dictated a certain level of restriction on the end format of the linguistic content database. For example, if content is to be reusable and transferrable, then the metatagging or database format (e.g. SQL, Oracle, etc.) to be used would need to be more generic

rather than specific (to allow for the addition of other languages or localised settings (LSs), for example). Similarly, the scale of the database content could be small (as in this instance) or large (if the content were to span more than two levels on the CEFRL for example or content for an entire series of textbooks), so the chosen solution would need to allow for scalability.

Databases are an electronic way to store information. Their makeup and nature should result in the easy access to information stored within them. There are many standard database formats and systems, from the widely used Microsoft Excel and Access, to more specialised SQL and Oracle formats and systems. The requirement in this instance was for a smaller scale, flexible format which could be applied to any number of technologies. Decoo (2011) reinforced the importance of retaining learning data in "easily transferrable formats, tied to universally used programs, rather than becoming totally dependent on tools that become obsolete or unusable in new operating systems" (p. 178).

Extensible Markup language (XML) is a markup language which describes data. It is a flexible format through which the designer composes their own labels to describe the data contained within them. Ward (2002) described XML technologies as being reusable in three ways – in the data processing engine that processes the files (when used on files with the same tags), in the file structure (how to structure and tag language learning content) and the file content. She maintained that XML allows a strict separation of data from processing mechanism (Ward, 2002). XML was a relatively new development when Ward (2002) reported on its use, but it has become the norm for web-based data feeds like RSS and transmitting information online. XML is an open standard, which is software- and hardware-independent.

Some practitioners have avoided using XML as its contents can become long and take a long time to parse and access. Colpaert (2004) described some of the limitations of XML when he noted how it does "not match all the requirements regarding functionalities on the level of natural language structure, content presentation and linguistic-didactic interaction" (p. 254). He argued that XML-based systems are "application dependent" as the tags are composed by the developer and specific to their original function and that XML is used mainly for "web-based technologies" (p. 285). However, one of the advantages of using XML is that it is platform neutral and extendable. More tags can be added at any stage to define any new data to be added to the database. Furthermore, applications can be tailored to the tag set, which would negate the assertion that XML is application dependent. Any database interface will be dependent on the format and structure of the database it is accessing.

XML was used as the database markup language for the linguistic content database. Aside from the advantages cited above, it also allowed intuitive tags to be composed and used, which

addressed a further caution from Braun (2005) regarding the data format and annotation of corpora being problematic for use in language learning. She reported that larger corpora with complex annotation schemes can have a listing of over 150 types among their POS tags. She pointed out that some of these tagsets are "too specific for the non-linguist" (Braun, 2005, p. 50). XML databases follow a hierarchical model, as the tags are nested. This sits well with the hierarchical nature of the PSC, which can be reflected across the sections of the raw linguistic content. The advantages of XML in its transferability were seen to outweigh the disadvantages of time in parsing/traversing. Koller (2007) outlined how there are two ways to parse an XML database – one which traverses the content in a stream (stream-based), thus overriding the parsing time challenges outlined by Colpaert (2004) or the other which parses the full XML file or tree (tree-based), before extracting the relevant data.

Colpaert (2004) proposed a three tier architecture which accounted for the database (Tier C), the teacher or administrator's interface (Tier B) and the learner's interface (Tier A) (as discussed in Chapter 3). Only Tier C is relevant to the discussion in this section, with Tiers B and C being relevant to the discussion in Section 7.1.4. Figure 7.1 illustrates Colpaert's architecture.

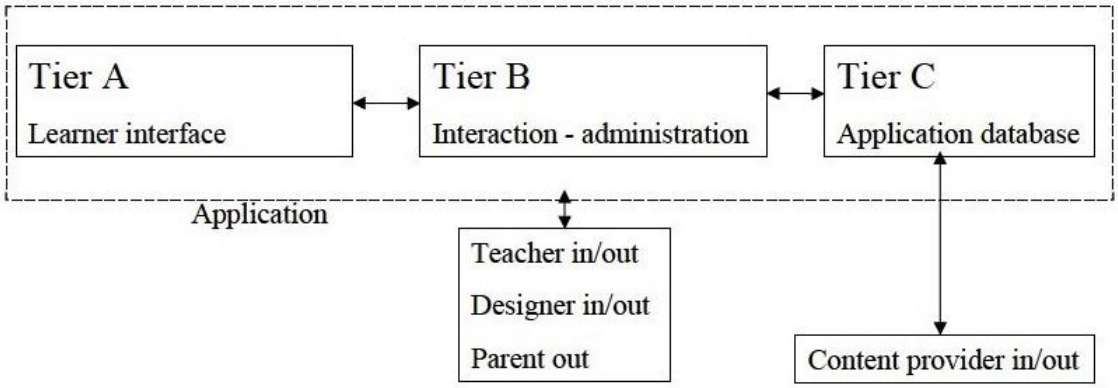


Figure 7.1 Colpaert's (2004) three tier architecture for specification of a solution within the RBRO-Model

7.1.2 Consolidating and converting LSs and raw linguistic content into a reusable electronic format

All of the data gathered during school-based research was housed in MS Word and Excel. Any data recorded orally or on paper was transferred into Excel for analysis. The relevant data and content from the localised settings (LSs) and raw linguistic content were drawn together and examined to ascertain the types of data present (in order to be able to best describe it through the XML tags) and the data which should be included or excluded from the database. Table 4.18 below which was also outlined in Chapter 4, summarises the data available and the data types.

When consideration was given to the types of data to be included or excluded from the linguistic content database, the deciding factor was the *usefulness of the information* to a teacher in the classroom or to the teacher's learners through the resource created. Nearly all of the datatypes were deemed relevant. The only datatype which was deemed to be too descriptive and not directly applicable to the teacher or learner in the classroom was the full listing of learner errors by type and frequency. The distilled top 5 errors and combined language awareness and error notes provided all of the salient information which was directly applicable to the local language learning context.

Data		**Data Type(s)**
Raw linguistic content		1. Headings (English) 2. Words (German) 3. Vocabulary items (Article Noun pairs) (German) 4. Full Sentences (questions, statements) (German) 5. POS data for all German content (2-4 as listed; tagset) 6. English and Irish translations of the German content (2-4 as listed; without POS details)
Localised Setting (LS) (2):	Linkage	References to internal section links (by heading name) (English)
	Integration	References to Subject>Strand>Strand unit of the PSC (English/Irish)
LS (4): Language Awareness		▪ Sentences referring to raw linguistic content sections (English/Irish) ▪ Sentences referring to relevant themes which are common to the raw linguistic content (e.g. dialects) (English/Irish) ▪ Linguistic/grammar items (e.g. compound words) (English/Irish) ▪ Language learning items (English)
LS (5): Error Analysis		▪ Listing of errors made by type and frequency for each raw linguistic content section ▪ Top 5 errors made in any one section as focus of potential errors for language learners ▪ Integrated language awareness and learner error focus notes, which help to identify pressure points in learning which can be addressed through teaching.

Table 7.1 Data types of relevance to the linguistic content database

The raw linguistic content comprised headings, German language (words, vocabulary items and full sentences), POS data attached to the German language and English and Irish translations of the German content (without POS details). Three levels of German language was deemed to be too many, so the XML structure was designed to incorporate the words and vocabulary items into the <multi_lang_word> element, and full sentences were represented by the element <multi_lang_sent>. Each of these elements allowed for the inclusion of POS data within the German content, and full sentence translations in any other language (to allow for English and Irish translations). The two data types were distinguished to account for partial sentences (article and noun pairs, or lists of adjectives) and full sentences. The challenges surrounding the partial

sentences were described in Chapter 6 when the raw linguistic content was being analysed and characterised. The distinction between the two data types ensured that partial sentences could be extracted and presented as such (e.g. lexicon entries), while full sentences would be more applicable to other types of resource outputs (e.g. handout). Further enhancement to the raw linguistic content itself would be the inclusion of POS data for the translation languages (English and Irish in this instance). This would allow a greater level of alignment between the three languages, but would also add complexity through the alignment process itself.

The data relating to the LS on linkage was all links within the German raw linguistic content sections themselves. The localised setting XML element (<Localised_setting>) comprised <note_type> and <note_detail>. "German section link" was used in the former for all linkage notes, while the section header was used as the corresponding <note_detail>. The LS integration details contained references to the Primary School Curriculum. Their inclusion mirrored the format outlined for integration, with the "PSC reference" being used as the <note_type> information.

Language awareness presented four different types of information as described in Table 4.18. Sub-themes were created to allow the easier access to the data contained within this LS note. These sub-themes fit within the generic XML structure which will be described in Section 7.1.3 below. The four language awareness 'types' which were applied to the <note_type> XML element are outlined in Table 7.2.

Type	**Description and example**	**XML entry**
Direct comparison	Comparing language in the section across the known language in English and Irish (e.g. Examine 'hello' expressions in German, English and Irish. Note 'dia duit and its responses)	Comp
Themes:	Focusing on language features (e.g. dialects, language families)	Theme
Grammatical	Examining areas of grammar (e.g. Word order, parts of speech, noun gender)	Gram
Language learning skills	Highlighting opportunities to draw on section content to teach language learning skills (e.g. examining cognates).	Skill

Table 7.2 Language awareness types applied to the generic <note_type> XML element

The data gathered around the LS on learner errors and its analysis revealed the three data types outlined in Table 4.18. As mentioned above, the first of these three data types, listing all errors made by type and frequency in each section, was deemed to be inappropriate for the teacher and learner in the classroom. The second of the data types provided a better distillation of this data through the top 5 errors encountered in any one section. Sub-themes were also created for the

error data to reflect the data types gathered and generated. The two error 'types' which were applied to the generic <note_type> XML element are described in Table 7.3.

Type	Description and example	XML entry
Top 5 errors	The top 5 errors made in any one section	Top 5
Integrated notes	Integrated language awareness and typical error notes	Integ

Table 7.3 Error types applied to the generic <note_type> XML element

XML was used so that the data could remain strictly separate from processing and to facilitate the addition of other languages later on by maintaining a modular approach in the data format. The relationships between the different data types was considered and included within the structure of the XML database. For example, all content was defined and structured under the <section> element, with this used as the uppermost tier of content data (aside from the root XML node element called <database>). The XML Document Type Definition (DTD) was drawn up to account for all of the data types as described. The database contents and structure are specified in more detail in Section 7.1.3 below.

7.1.3 Specifying the database content

A Document Type Definition (DTD) was used to specify the contents and structure of the XML linguistic content database. There was no requirement to check the content of any of the XML elements (e.g. date format conforming to DDMMYY), resulting in a DTD being more suited to the definition than an XML schema.

Maintaining a generic structure for transferability

The use of attributes within the XML database was kept to a minimum so that the structure would not be confined to a limited set of attribute values. For example, rather than defining a set of three language attributes in the sentence element as being German, English and Irish (e.g. <sentence lang="Eng">), the additional elements of <sentence_language> and/or <word_language> were used, so that any language could be inserted to represent the content of the parent element of <sentence> or <word>. This increased the length of the database file, and added to the parse complexity, but it did ensure that the database structure and labelling remained generic allowing the insertion of any language.

Optional multimedia file references can sit alongside any sentence or word elements in the database. Their inclusion is optional, and the generic tags used, ensure that any type of multimedia file(s) can be added to the content. The DTD requires that the 'type' and the 'file location/name' are specified as follows

… <mm_file>

```
        <mm_type>image</mm_type>
        <mm_name>image.jpg</mm_name>
    </mm_file>
```

The XML tagset used around the localised settings also ensured a more generic construct. The LS names were not specified to allow for any LS to be added to the linguistic content database within the <LS_name> element. Similarly, the notes attached to each of the LS elements were structured with a <note_type> element to allow the content to be entered in each case.

Attributes were used in one instance where the POS information was inserted. Attributes are especially useful, where the detail contained within them acts as a reference point for the element. The use of POS data was as an identifier for the selection and pinpointing of information. It could also be used as a piece of information in itself, in database outputs such as lexicon entries. The TreeTagger POS tags (see Appendix M) were used as attributes of the <word> element, for example,

```
        <word pos="ART">der</word>
        <word pos="N">Mann</word>
```

The POS attribute for the database content is an optional attribute, which can be omitted. Other local contexts may not require the addition of POS tags. If POS tags were relevant to another local context, and the tagset was different, an update to the XML DTD would allow the new tagset to be applied. At the same time, the TreeTagger tool is a language independent tool and could readily be used in the tagging of new and relevant data for alternative local contexts. This solution was chosen in this instance in contrast to those described above, as the additional insertion of a POS label element would have substantially increased the length of the database file (every German word in the database has associated POS data). Some small changes to the POS tags were made in three instances to allow for the easier parsing of data - $., $, and $(were changed to FULL, COMMA and BRACKET respectively.

Figure 7.3 shows an excerpt from the XML linguistic content database which illustrates the generic labelling for language content, and the use of attributes to house POS data. A full section excerpt from the linguistic content database for *Section One: Greetings* can be found in Appendix S

```
<multi_lang_sent>
        <sent>
                <sent_lang>German</sent_lang>
                <word pos="PPER">Mir</word>
                <word pos="VVFIN">geht</word>
                <word pos="PPER">'s</word>
                <word pos="ADJD">gut</word>
                <word pos="$.">.</word>
        </sent>
        <sent>
                <sent_lang>English</sent_lang>
                <full_sent>I'm good.</full_sent>
        </sent>
        <sent>
                <sent_lang>Irish</sent_lang>
                <full_sent>Táim go brea.</full_sent>
        </sent>
</multi_lang_sent>
```

Figure 7.2 Extract from XML linguistic content database (Section One: Greetings)

Structure of the XML linguistic content database

The XML tree of Figure 7.3 illustrates the format which is defined in the DTD below. The only missing information relates to the number of allowable occurrences of each of the elements contained within the database. Most elements occur one or more times, with the exceptions of the <mm_file> element (used to house multimedia information) and the <Localised_setting> element (used to house the LS data), which can occur zero or more times.

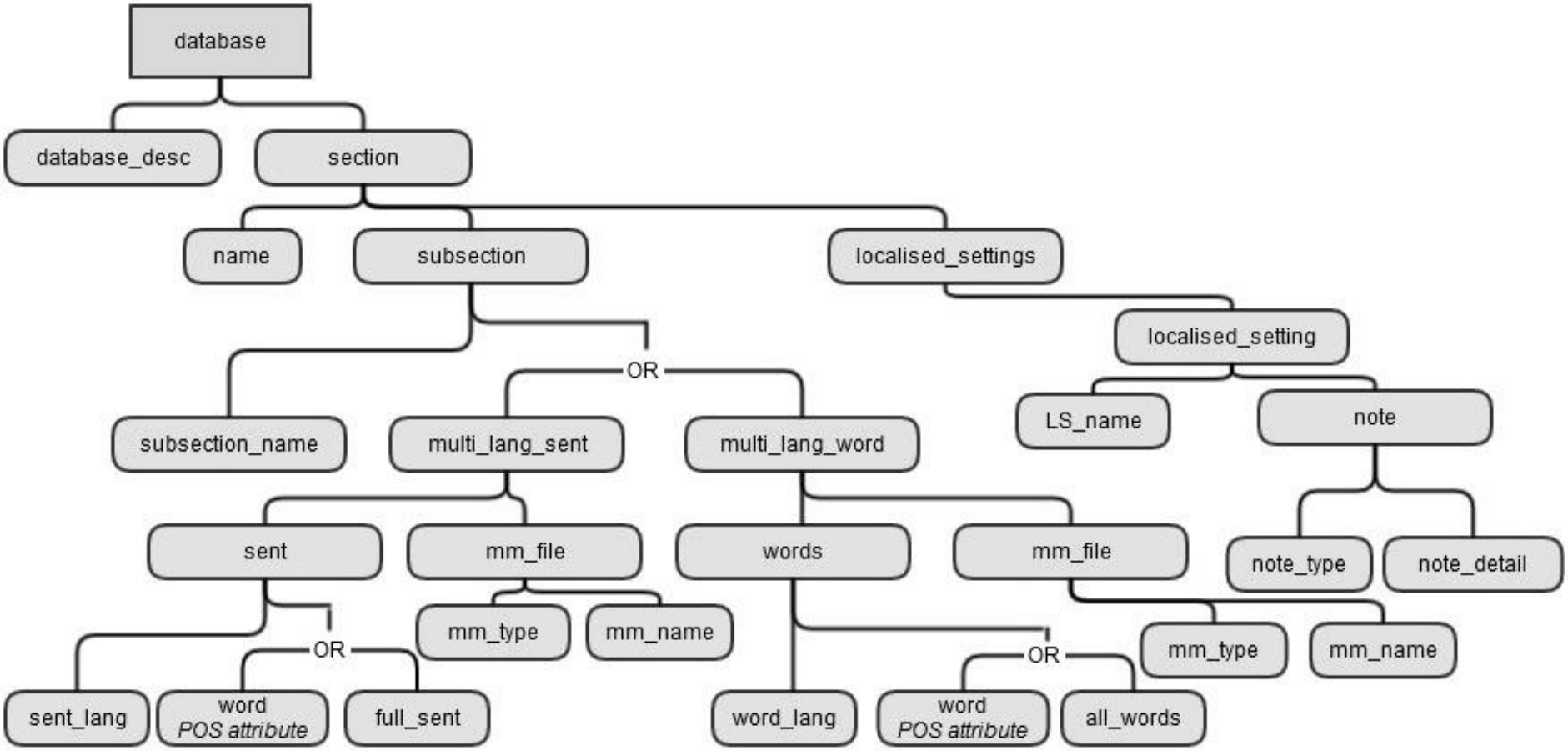


Figure 7.3 XML tree representing the structure of the linguistic content database

The DTD for the linguistic content database is provided in Figure 7.4 below. It was used to validate the structure of the XML document, as well as to describe the structure of the linguistic content database and dependency of the XML elements on one another (e.g. child node, sister node etc.).

```
<!ELEMENT database (database_desc, section+)>
<!ELEMENT database_desc (#PCDATA)>

<!ELEMENT section (name, subsection+, Localised_settings*)>
<!ELEMENT name (#PCDATA)>

<!ELEMENT subsection (subsection_name, (multi_lang_sent|multi_lang_word)+)>
<!ELEMENT subsection_name (#PCDATA)>

<!ELEMENT multi_lang_sent (sent+, mm_file*)>
<!ELEMENT sent (sent_lang, (word+|full_sent))>
<!ELEMENT sent_lang (#PCDATA)>
<!ELEMENT full_sent (#PCDATA)>
<!ELEMENT word (#PCDATA)>
<!ATTLIST word pos
(ADJA|ADJD|ADV|APPR|APPRART|APPO|APZR|ART|CARD|FM|ITJ|KON|KOKOM|KOUI|KOUS|NE
|NN|PAV|PDAT|PDS|PIAT|PIDAT|PIS|PPER|PRF|PPOSS|PPOSAT|PRELAT|PRELS|PTKA|PTKA
NT|PTKNEG|PTKVZ|PTKZU|PWS|PWAT|PWAV|TRUNC|VAFIN|VAIMP|VAINF|VAPP|VMFIN|VMINF
|VMPP|VVFIN|VVIMP|VVINF|VVIZU|VVPP|XY|FULL|COMMA|BRACKET) #IMPLIED>
<!ELEMENT image_file (#PCDATA)>

<!ELEMENT mm_file (mm_type, mm_name)>
<!ELEMENT mm_type (#PCDATA)>
<!ELEMENT mm_name (#PCDATA)>

<!ELEMENT multi_lang_word (words+, mm_file*)>
<!ELEMENT words (word_lang, (word+|all_words))>
<!ELEMENT all_words (#PCDATA)>
<!ELEMENT word_lang (#PCDATA)>


<!ELEMENT Localised_settings (Localised_setting+)>
<!ELEMENT Localised_setting (LS_name, note+)>
<!ELEMENT LS_name (#PCDATA)>
<!ELEMENT note (note_type, note_detail+)>
<!ELEMENT note_type (#PCDATA)>
<!ELEMENT note_detail (#PCDATA)>

```

```
<!ELEMENT database (database_desc, section+)>
<!ELEMENT database_desc (#PCDATA)>

<!ELEMENT section (name, subsection+, Localised_settings*)>
<!ELEMENT name (#PCDATA)>

<!ELEMENT subsection (subsection_name, (multi_lang_sent|multi_lang_word)+)>
<!ELEMENT subsection_name (#PCDATA)>

<!ELEMENT multi_lang_sent (sent+, mm_file*)>
<!ELEMENT sent (sent_lang, (word+|full_sent))>
<!ELEMENT sent_lang (#PCDATA)>
<!ELEMENT full_sent (#PCDATA)>
<!ELEMENT word (#PCDATA)>
<!ATTLIST word pos
(ADJA|ADJD|ADV|APPR|APPRART|APPO|APZR|ART|CARD|FM|ITJ|KON|KOKOM|KOUI|KOUS|NE
|NN|PAV|PDAT|PDS|PIAT|PIDAT|PIS|PPER|PRF|PPOSS|PPOSAT|PRELAT|PRELS|PTKA|PTKA
NT|PTKNEG|PTKVZ|PTKZU|PWS|PWAT|PWAV|TRUNC|VAFIN|VAIMP|VAINF|VAPP|VMFIN|VMINF
|VMPP|VVFIN|VVIMP|VVINF|VVIZU|VVPP|XY|FULL|COMMA|BRACKET) #IMPLIED>
<!ELEMENT image_file (#PCDATA)>

<!ELEMENT mm_file (mm_type, mm_name)>
<!ELEMENT mm_type (#PCDATA)>
<!ELEMENT mm_name (#PCDATA)>

<!ELEMENT multi_lang_word (words+, mm_file*)>
<!ELEMENT words (word_lang, (word+|all_words))>
<!ELEMENT all_words (#PCDATA)>
<!ELEMENT word_lang (#PCDATA)>


<!ELEMENT Localised_settings (Localised_setting+)>
<!ELEMENT Localised_setting (LS_name, note+)>
<!ELEMENT LS_name (#PCDATA)>
<!ELEMENT note (note_type, note_detail+)>
<!ELEMENT note_type (#PCDATA)>
<!ELEMENT note_detail (#PCDATA)>

```

Figure 7.4 Document-Type Definition (DTD) for the XML linguistic content database

7.1.4 Assessing conformance to the database requirements

The four criteria for database development outlined in Section 7.1.1 are each briefly considered below for the database fit to their requirement.

1. *be accessible to the end user for the quick and easy development of resources*

The teacher is the end 'user' who develops each of the proof of concept resources from the teacher interface with the linguistic content database. A series of radio-buttons were used to allow the content to be accessed in a 'quick and easy' fashion, the end result of which is the narrowing down of applicable content from the linguistic content database for use in one of the three resources described below.

2. *account for the localisation of resources through its structure and content*

The XML linguistic content database was developed so that the markup describes the content held in the database as well as structures it. At the same time, the labels used needed to be generic

enough to allow for the addition of alternative LSs to account for localisation in other local contexts. The data types listed in Section 7.1.2 account for the varying localisation factors which were considered in the development of the raw linguistic resources – curriculum-based raw linguistic resources, and localised settings which all account for the local research context of the primary school German language learning classroom in Ireland. Through the generic metatags used and described in Section 7.1.3, scope is provided for using the same structure for alternative local contexts.

3. *allow for reusability and transferability of content /*
4. *ensure the separation of content from its product*

The use of an XML database for storing the content of the raw linguistic content and associated localised settings ensured a separation of content from its end product. The linguistic content database holds language learning content only, which can be accessed and used in the generation of language teaching and learning resources (see below for examples). Content can be reused in many ways through the development of alternative output templates or resources which all draw on the same database content.

Content can be transferred to alternative settings through the inclusion of localised settings. These are settings which can be switched on or off, accessed or hidden, depending on their relevance to other related contexts. They can similarly be used as a basis for the generation of related content, as seen in Section 4.1.3 where the example of the use of the German raw linguistic content was reused as a basis for its equivalent content in French.

7.2 Examples of the linguistic content database in use

Three proof of concept outputs were developed to illustrate how the content of the linguistic content database could be used to generate alternative forms of teaching and learning resources. This represents Phase IV of work, as outlined in Chapter 4. The three resource types were (1) teacher report, (2) classroom handout, and (3) lexicon creation. Each of these outputs will be discussed and illustrated across the coming sections.

Producing teaching and learning outputs from databases of content is not an unusual activity. Section 3.2.5 outlined how Colpaert (2004), Decoo and Colpaert (1997), Knapp (2004) and Ward (2002) have all generated and used content databases in the creation of language teaching and learning resources. Schonenberg and Heughebaert (1996) similarly outlined how they had used a content source to not only produce interactive exercises, but also develop a textbook.

They described that usually, this process is reversed, where form to content is the order of development (textbook to content). However, they argued that the content to form order is also a rational and proven path.

The sample outputs illustrated in the coming sections are proof of concept examples of how the linguistic content database can be employed and used by the teacher to create different resource types for his/her learners. While the outputs are functional and operate effectively, the design and look of the outputs is far from an attractively designed textbook or website. Similarly, the extent of user selection on the database could be further extended to allow for more fine-grained access to the content held within the linguistic content database, for example, selection of the relevant languages for the output (German/Irish in Irish-medium schools), selection of particular words for lexicon creation and mixing content from a variety of sections. The purpose of these outputs was not to produce a final end product, rather illustrate the *potential* application of the core area of research in this work - the linguistic content database. For the same reason, these proof of concept outputs were not tested with teachers or publishing houses to ascertain their suitability as end products. As Figure 7.5 illustrates below, the development of template-based outputs for these end-products would require customisation by a developer for the individual end user and their resource requirement (be it a teacher requiring a handout, or a publishing house accessing the content in the development of a textbook).

HTML, XML, PHP and DTD were the technologies used to implement the proof of concept outputs. They all revolve around taking in user selection of required data (HTML, PHP), selecting the required data from the XML database (PHP) and outputting the relevant data into the resource type (PHP). Figure 7.5 outlines how three proof of concept outputs were developed and accessed from one home screen menu. It illustrates the flow of information between the linguistic content database and the resource output screens.

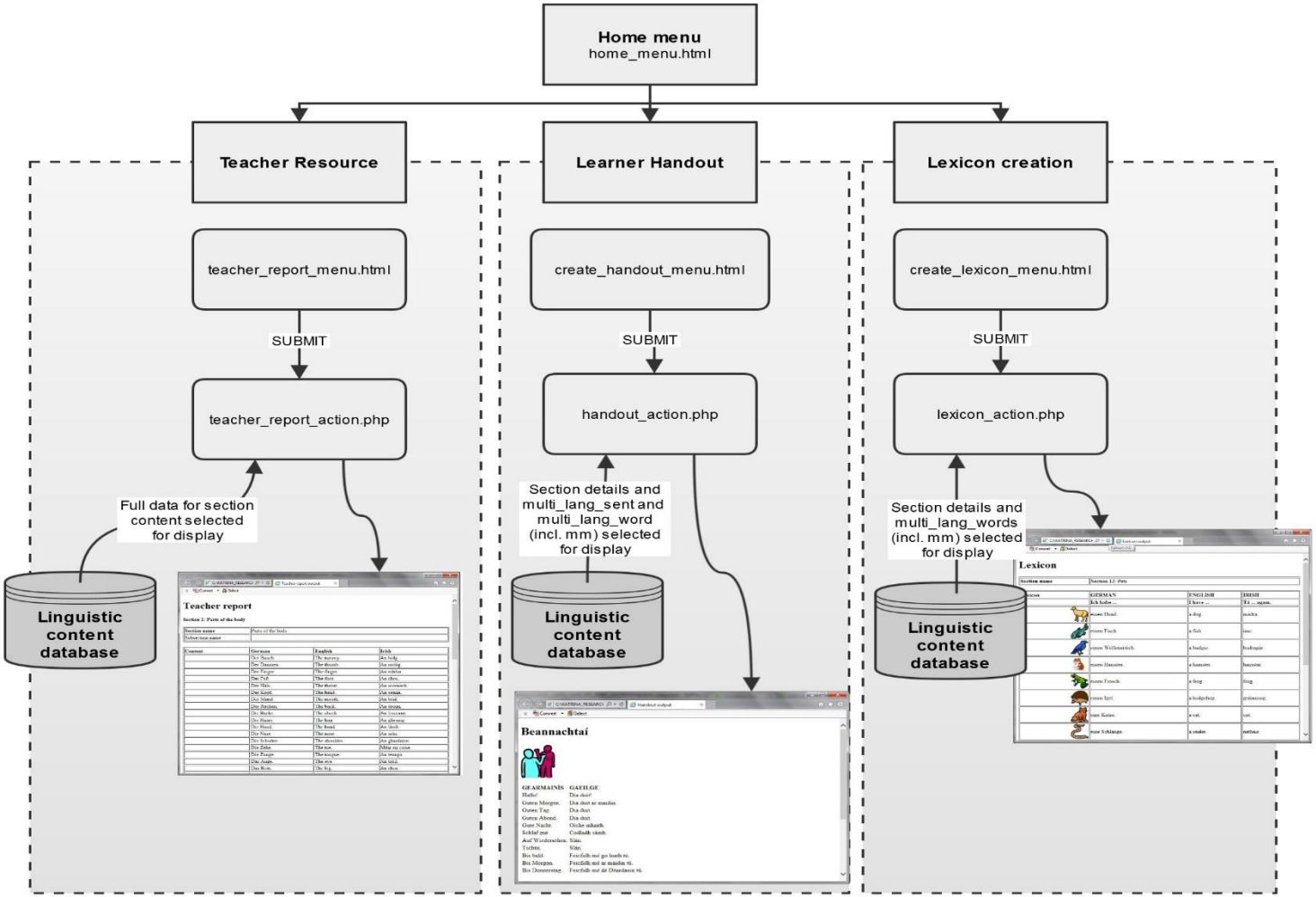


Figure 7.5 Flow of information between the proof of concept output modules and the linguistic content database

One home screen was used as the menu for the three proof of concept outputs. The home screen is illustrated in Figure 7.6 and leads on to the proof of concept output interface which is selected.

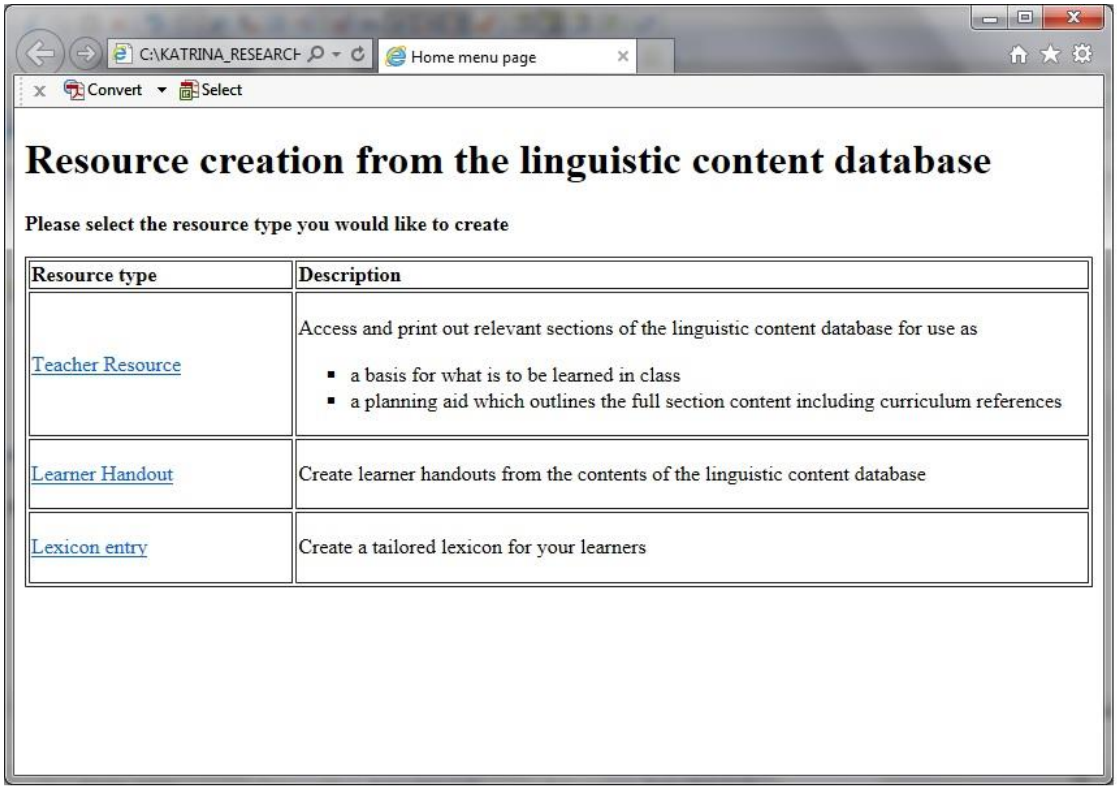


Figure 7.6 Menu page for the selection of proof of concept resource type

7.2.1 Teacher report

Teachers in Ireland were reported to rely too heavily on ready-made resources such as textbooks to guide their planning and teaching (DES, 2007; Gilleece, et al, 2012; NCCA, 2005a, 2008a, 2010). With this challenge in mind, the teacher report proof of concept output was devised to support teachers in their curriculum-based planning and reporting.

Teachers in Ireland are required to produce a long-term plan to account for yearly (usually established teachers) or termly (usually newly-qualified teachers) teaching and learning, as well an associated short-term plan which accounts for weekly progress. They are also required to produce a plan called the Cuntas Míosúil / *monthly progress report* to indicate what has been taught and learned during any calendar month. The teacher report allows teachers to access and print out relevant sections of the linguistic content database. This output could be used by teachers as a means of accessing the full or partial content of the linguistic content database (section by section), as a basis for what is to be learned in class or as a planning aid which outlines the full section content from which they can pick and choose relevant topics, areas of integration, areas of linkage or areas of Language Awareness.

The sequence of the teacher accessing the data of the linguistic content database to tailor the output to his/her needs is outlined across Figure 7.7 to Figure 7.10. The initial process of selecting the relevant section of the linguistic content database is illustrated in Figure 7.7. The remaining three figures show how the relevant content is called from the linguistic content database and displayed in table format (Figure 7.8, Figure 7.9 and Figure 7.10).

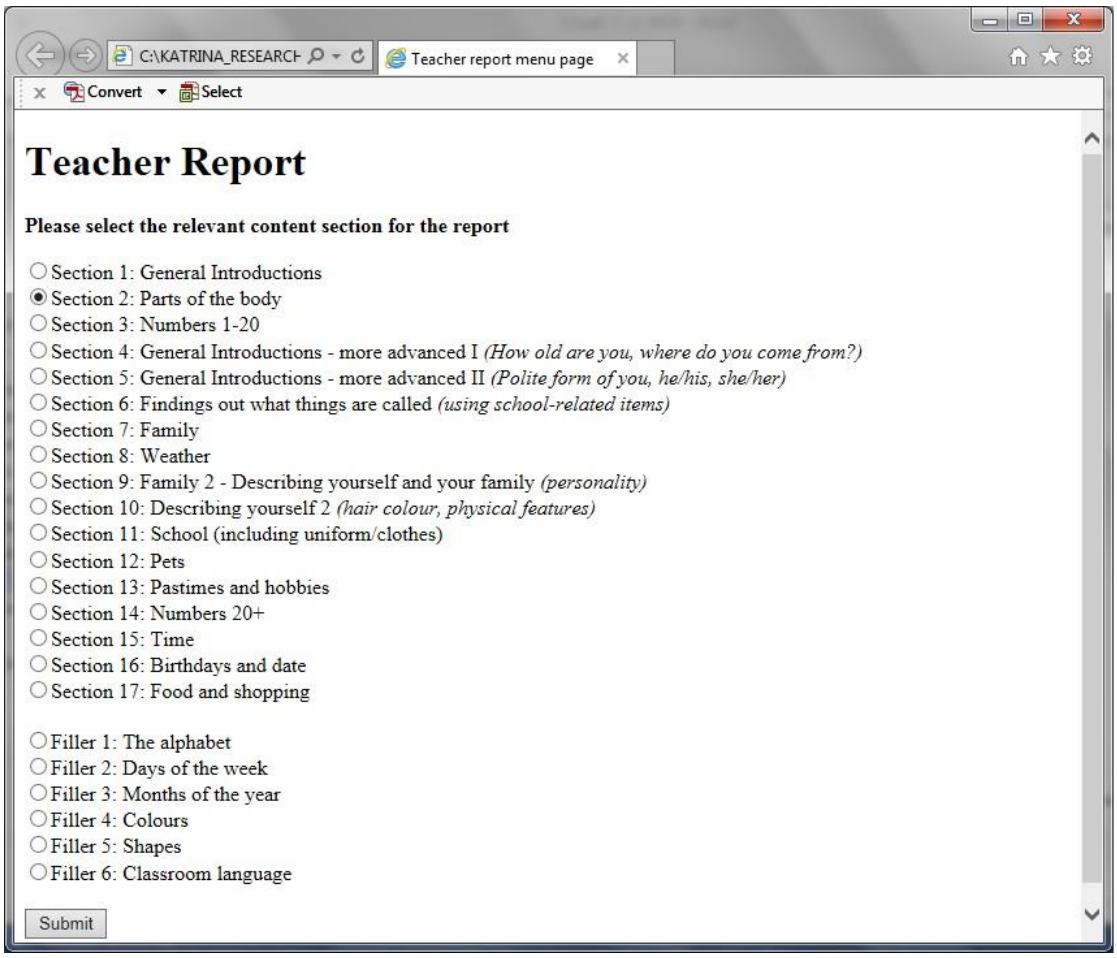
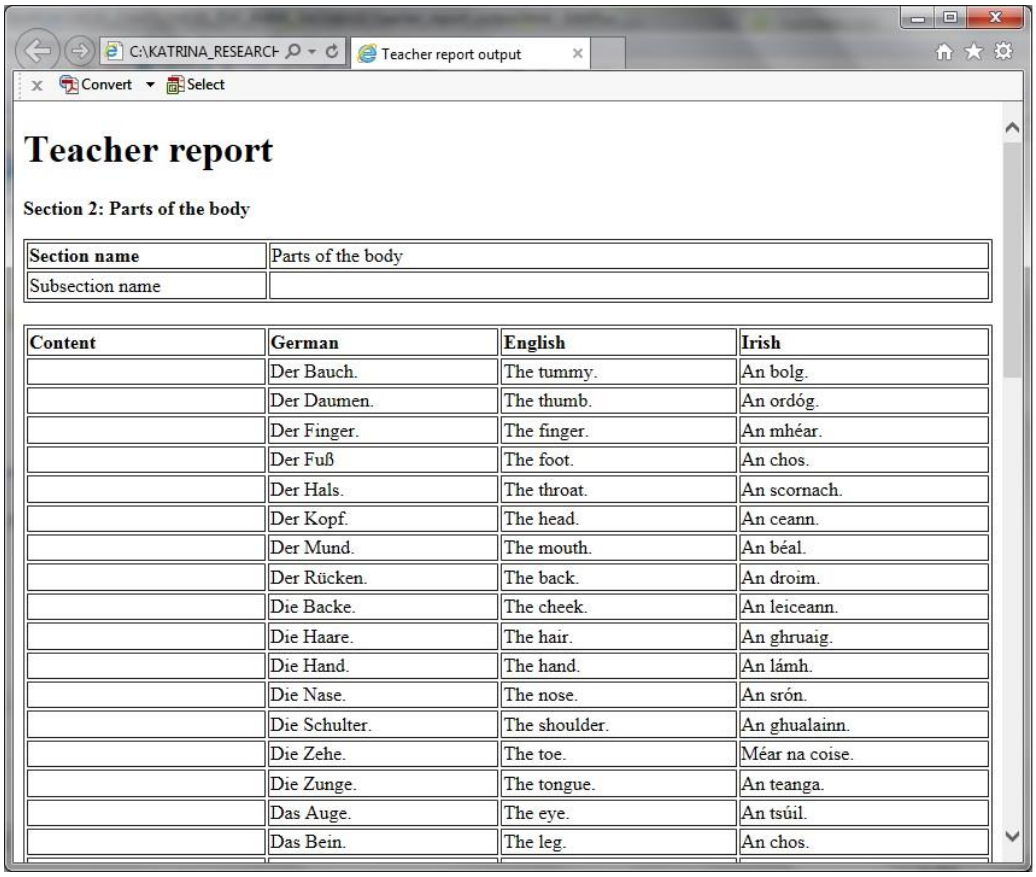


Figure 7.7 Section selection within teacher report



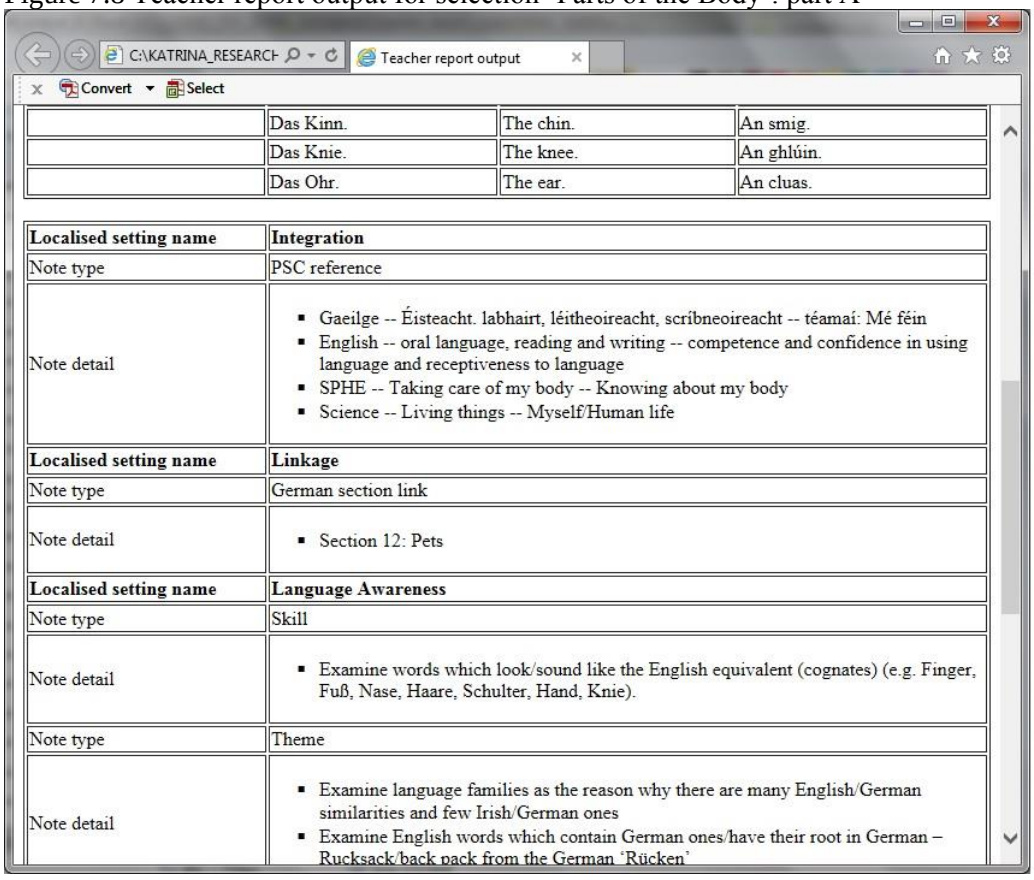
Teacher report

Section 2: Parts of the body

Section name	Parts of the body
Subsection name	

Content	**German**	**English**	**Irish**
	Der Bauch.	The tummy.	An bolg.
	Der Daumen.	The thumb.	An ordóg.
	Der Finger.	The finger.	An mhéar.
	Der Fuß	The foot.	An chos.
	Der Hals.	The throat.	An scornach.
	Der Kopf.	The head.	An ceann.
	Der Mund.	The mouth.	An béal.
	Der Rücken.	The back.	An droim.
	Die Backe.	The cheek.	An leiceann.
	Die Haare.	The hair.	An ghruaig.
	Die Hand.	The hand.	An lámh.
	Die Nase.	The nose.	An srón.
	Die Schulter.	The shoulder.	An ghualainn.
	Die Zehe.	The toe.	Méar na coise.
	Die Zunge.	The tongue.	An teanga.
	Das Auge.	The eye.	An tsúil.
	Das Bein.	The leg.	An chos.

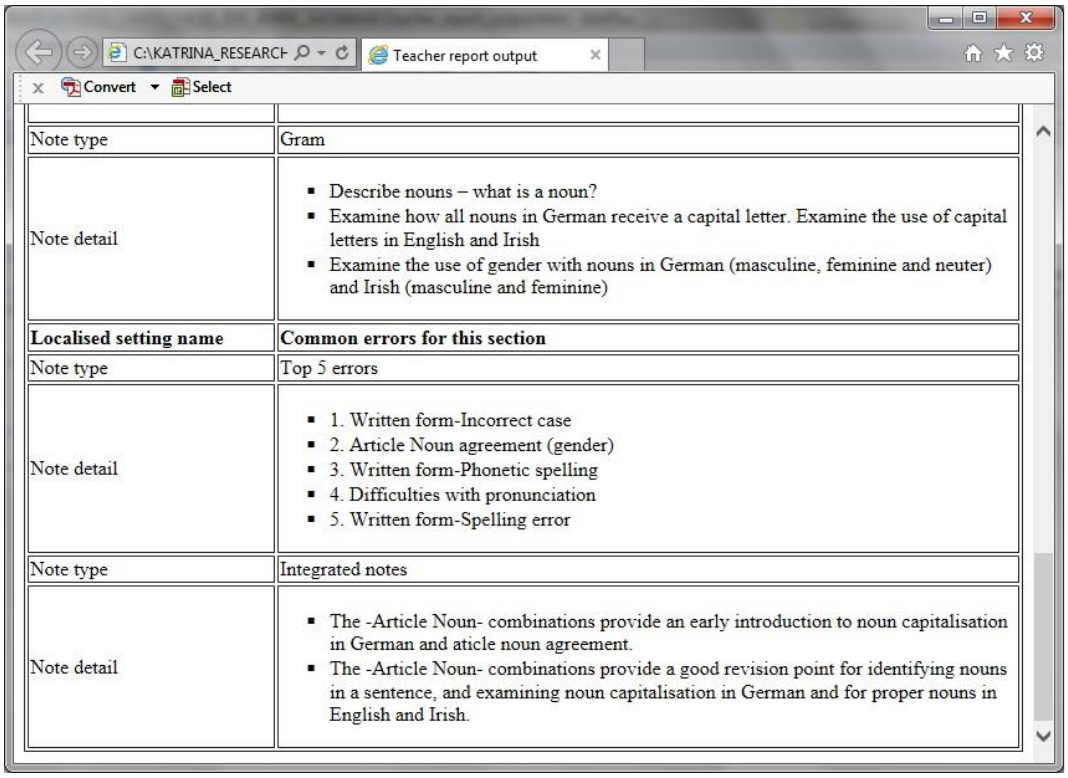
Figure 7.8 Teacher report output for selection ‘Parts of the Body’: part A



	Das Kinn.	The chin.	An smig.
	Das Knie.	The knee.	An ghlúin.
	Das Ohr.	The ear.	An cluas.

Localised setting name	**Integration**
Note type	PSC reference
Note detail	▪ Gaeilge -- Éisteacht. labhairt, léitheoireacht, scríbneoireacht -- téamaí: Mé féin ▪ English -- oral language, reading and writing -- competence and confidence in using language and receptiveness to language ▪ SPHE -- Taking care of my body -- Knowing about my body ▪ Science -- Living things -- Myself/Human life
Localised setting name	**Linkage**
Note type	German section link
Note detail	▪ Section 12: Pets
Localised setting name	**Language Awareness**
Note type	Skill
Note detail	▪ Examine words which look/sound like the English equivalent (cognates) (e.g. Finger, Fuß, Nase, Haare, Schulter, Hand, Knie).
Note type	Theme
Note detail	▪ Examine language families as the reason why there are many English/German similarities and few Irish/German ones ▪ Examine English words which contain German ones/have their root in German – Rucksack/back pack from the German ‘Rücken’

Figure 7.9 Teacher report output for selection ‘Parts of the Body’: part B



Note type	Gram
Note detail	▪ Describe nouns – what is a noun? ▪ Examine how all nouns in German receive a capital letter. Examine the use of capital letters in English and Irish ▪ Examine the use of gender with nouns in German (masculine, feminine and neuter) and Irish (masculine and feminine)
Localised setting name	**Common errors for this section**
Note type	Top 5 errors
Note detail	▪ 1. Written form-Incorrect case ▪ 2. Article Noun agreement (gender) ▪ 3. Written form-Phonetic spelling ▪ 4. Difficulties with pronunciation ▪ 5. Written form-Spelling error
Note type	Integrated notes
Note detail	▪ The -Article Noun- combinations provide an early introduction to noun capitalisation in German and aticle noun agreement. ▪ The -Article Noun- combinations provide a good revision point for identifying nouns in a sentence, and examining noun capitalisation in German and for proper nouns in English and Irish.

Figure 7.10 Teacher report output for selection ‘Parts of the Body’: part C

7.2.2 Handout generation

The raw linguistic content was presented to the children participating in the classroom-based research through handouts and worksheets. As a useful source of reference for language which is to be covered during lessons, handouts can also build up over time to create a semi-textbook for children. Even in textbook-dominant classrooms, supplementary handouts and worksheets are often introduced and used.

Using the linguistic content database as a source of content for the generation of handouts overcomes the challenges associated with textbook use, especially where they are more global in nature and require localisation (Prodromou and Mishan, 2008). In the same way, the use of the linguistic content database also combats the heretofore distance of published teaching and learning materials from their curriculum-basis (Harbison, 2008).

The figures across the following pages illustrate the process of creating a template-based handout for Section 1: Greetings. Figure 7.11 illustrates the section selection for the handout creation.

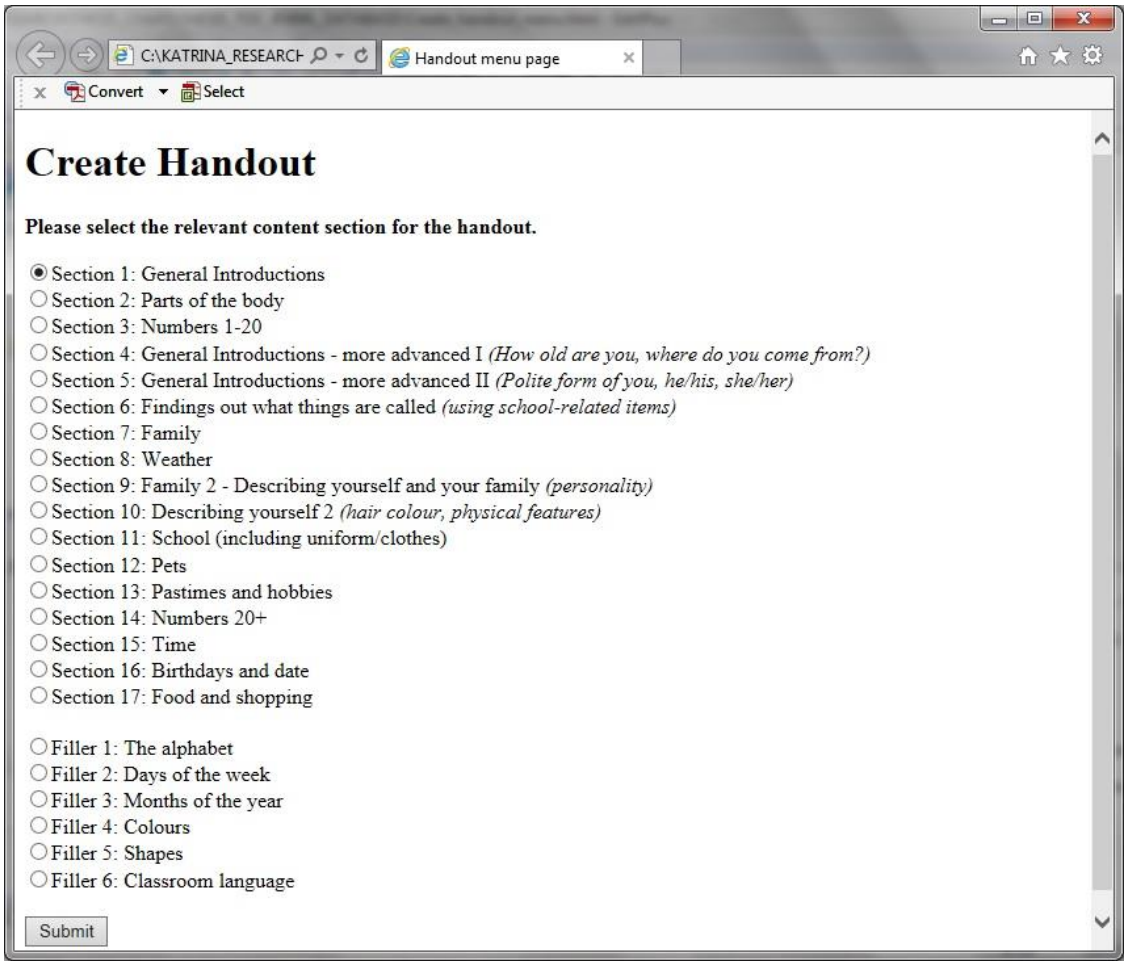


Figure 7.11 Section selection within handout creation

The handout template calls on the full sentences (the <multi_lang_sent> XML element) and the German and Irish language content. This proof of concept output shows how varying elements of the linguistic content database can be easily manipulated to create a localised resource. A handout such as this would be useful in an Irish-medium school, where teachers have reported a dearth of resources in the Irish language (NCCA, 2008a). Some image files were also added to the linguistic content database. Where present, they are included in the resource output. Figure 7.12 and Figure 7.13 show the output of the handout creation process.

The application of this handout resource in the classroom would follow the same use as any paper-based consumable resource. The handout content remains separated from its pedagogical focus, which is planned for and integrated by the teacher in the classroom. The handout provides a point of reference for the language to be taught and learned.

Figure 7.12 Output of the handout creation process for 'Greetings': part A

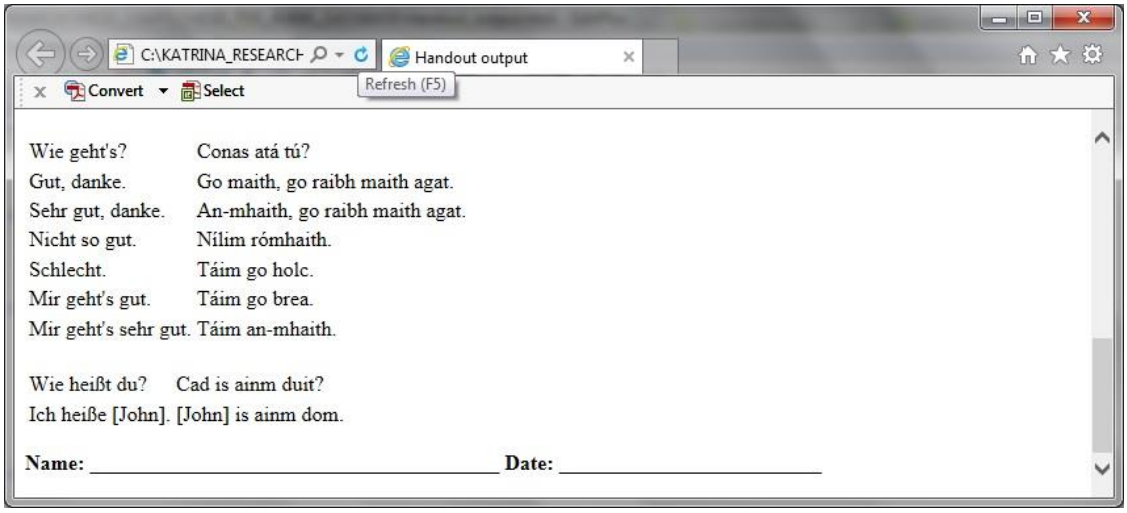


Figure 7.13 Output of the handout creation process for 'Greetings': part B

7.2.3 Create lexicon

Lexica (paper-based or digital) are useful tools in the classroom for supporting teaching and learning. They can be used to provide assistance for learners to traverse a comprehension text (glossary), to revise known or newly presented vocabulary or to track learner progress across language being learned (Decoo, 2011).

The linguistic content database is accessed by section, as for the previous two proof of concept outputs. Once the relevant section has been chosen by the teacher for inclusion in the lexicon,

all of the vocabulary items within that section (incomplete sentences labelled as <multi_lang_word>) are displayed onscreen. Their English and Irish translation is also displayed alongside them along with any relevant image files. An additional XML element within <multi_lang_word> was added to the linguistic content database to allow full sentences to be created from the words contained within the incomplete listing, where required. For example, the German words in 'pets' were in the accusative case (e.g. the article in '*einen Hund'* is in the accusative case and requires a nominative subject) and intended to be paired with the sentence '*I have a…*'. This missing sentential information was provided in all three languages across the linguistic content database. The requirement for this additional element illustrates a shortcoming in the database content, in that its language is confined to the scenarios through which it is presented, for example, art-noun pairs in the accusative case which form part of a section context. Further use of CL/NLP tools could assist in automatically providing alternative content to match the output requirements. They could similarly be integrated into the output process to deliver the output language in its optimum format, or case (e.g. nominative, accusative).

Figure 7.14 illustrates how the relevant section is chosen for the lexicon creation. The proceeding figures, Figure 7.15 and Figure 7.16 illustrate the lexicon output.

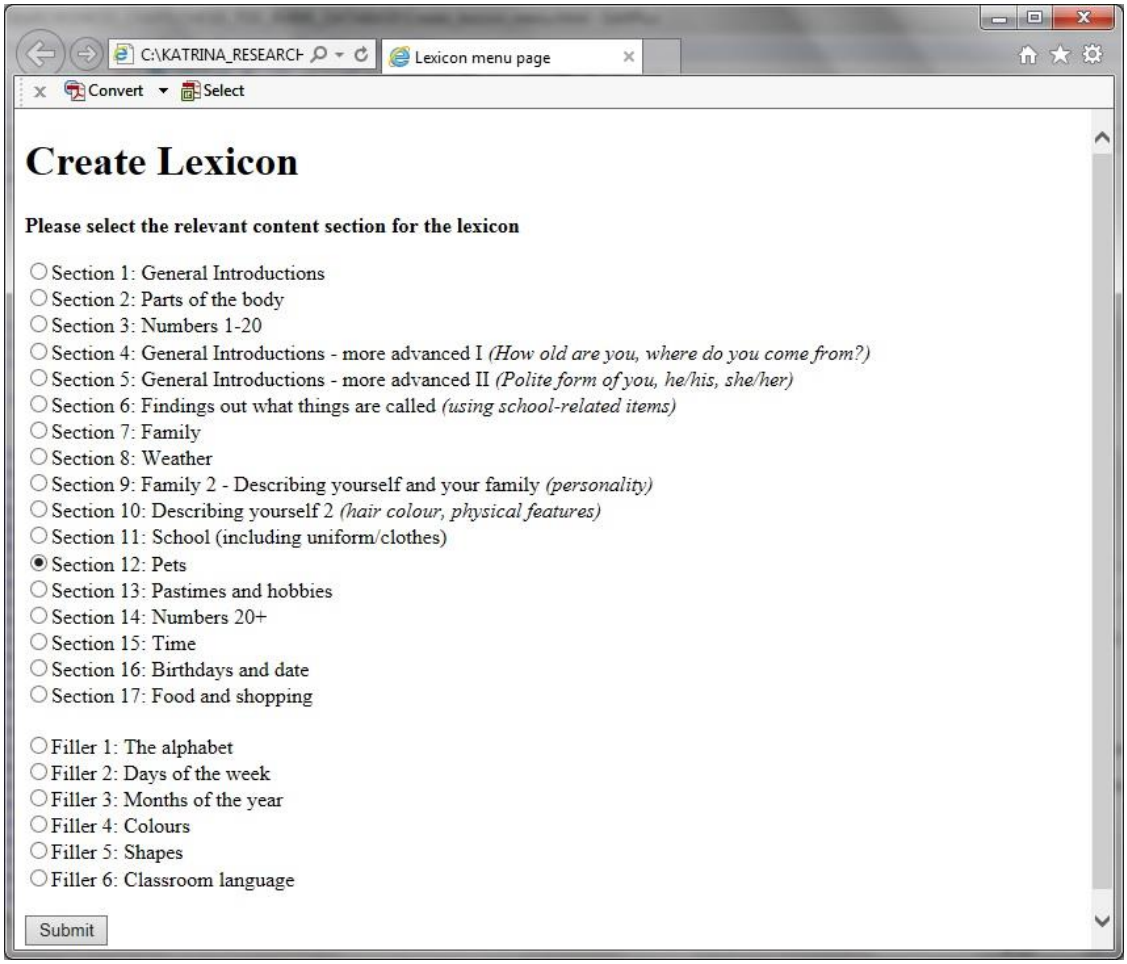
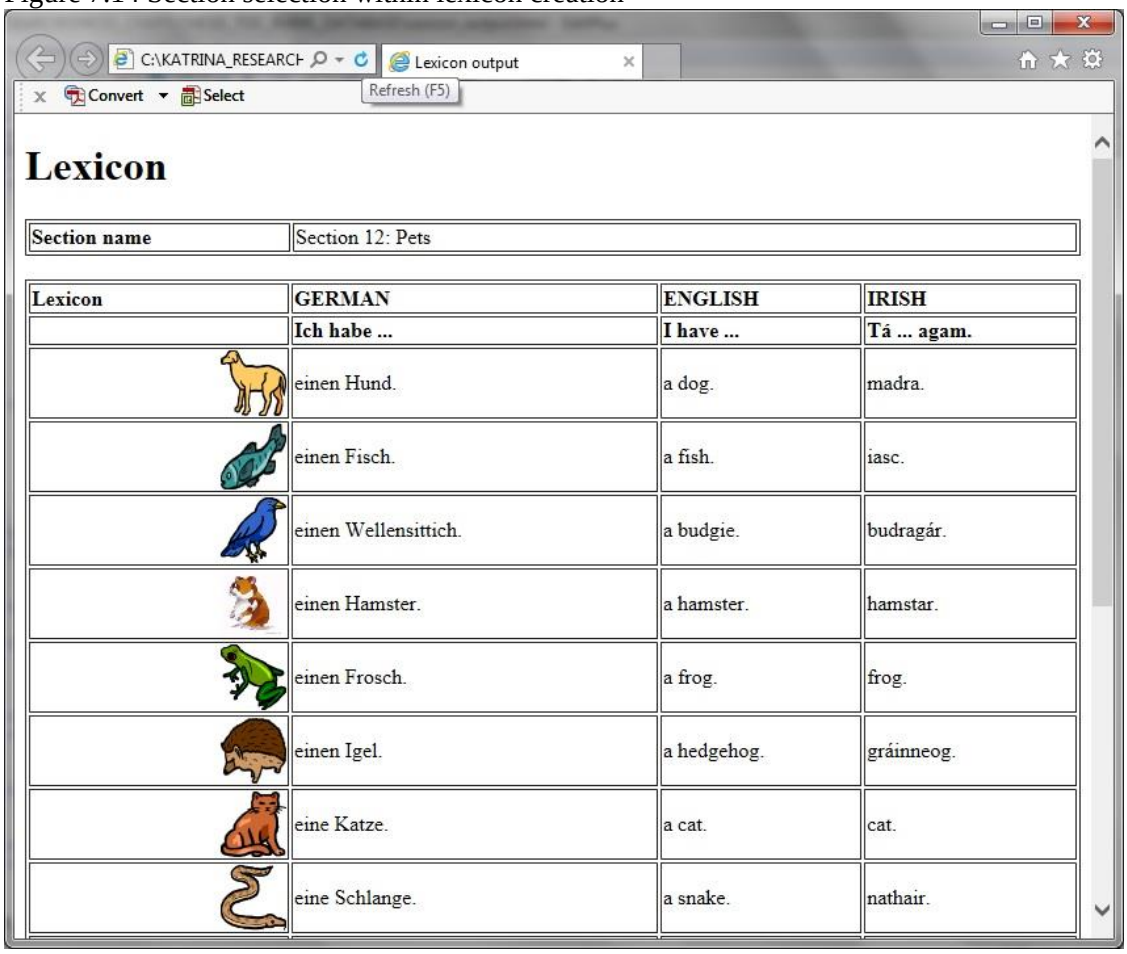


Figure 7.14 Section selection within lexicon creation



Lexicon

Section name	Section 12: Pets

Lexicon	GERMAN	ENGLISH	IRISH
	Ich habe ...	**I have ...**	**Tá ... agam.**
	einen Hund.	a dog.	madra.
	einen Fisch.	a fish.	iasc.
	einen Wellensittich.	a budgie.	budragár.
	einen Hamster.	a hamster.	hamstar.
	einen Frosch.	a frog.	frog.
	einen Igel.	a hedgehog.	gráinneog.
	eine Katze.	a cat.	cat.
	eine Schlange.	a snake.	nathair.

Figure 7.15 Lexicon output for selection ‘Pets’: part A

	eine Shildkröte.	a tortoise.	toirtís.
	eine Maus.	a mouse.	luchóg.
	eine Schnecke.	a snail.	seilide.
	eine Spinne.	a spider.	damhán alla.
	ein Kaninchen.	a rabbit.	coinín.
	ein Pferd..	a horse.	capall.
	ein Meerschweinchen.	guinea pig.	muc ghuine.

Figure 7.16 Lexicon output for selection ‘Pets’: part B

The lexicon proof of concept output could be advanced further by allowing the teacher to select the relevant lexicon entries (by word), or languages which he/she would like to use. The visual display of the lexicon would also benefit from being designed by a professional to improve the visual aesthetic of the output.

Lexica such as these, or selections of content, could also be used as the input to other types of language learning tools. One particular tool is the creation of vocabulary SMS messages for learners, as described in Keogh (2011) and NCCA (2010c). Similarly, the presence of the same vocabulary across three languages can form a useful language awareness tool. Anecdotal evidence gathered during time in the classroom revealed that children enjoyed this area of focus as it reinforced their existing knowledge of language, showing that they already knew a lot about language(s). They also enjoyed the comparison of what they did and didn't know across the languages, such as working out the grammatical 'systems' of plurals.

Content- and activity-based resources in the classroom

The three resource outputs presented are all content-based resources. Activity-based resources could also be created, such as gap-fill exercises which isolate particular parts of speech for filling, translation-based exercises which require the learners to translate content from one language into another language or matching exercises which require an image to be matched to its relevant word or sentence.

In the classroom, the handout could be used as a source of reference for learners to part-take in a task-based role-play where each learner needs to find out how other learners are doing ('*How are you?'*) or learn the fictitious names of 5 other classmates. A tailored lexicon which is based on classroom items could form the basis of a 'search and find' activity, where each item on the list must be found in the classroom, its name said out loud in the target language and then ticked off the list. A lexicon could also be used in a more traditional way of building up known vocabulary over time, with an activity centring on practising known words and sentences, and identifying new ones. With content abound from the linguistic content database, the scale and scope of the potential associated language learning activities would lie in the didactic professionalism and imagination of the teacher.

7.3 Summary and conclusion

This chapter outlined the processes involved in Phases III and IV of work. It described how the linguistic content database was developed using XML as the mark-up language. The cited advantages of XML were that it is platform neural, extendable and allowed for the development of customised and intuitive tags, while also remaining generic. An XML DTD outlined the

structure and relationships between the various elements of the XML database. The raw linguistic resources were consolidated along with their English and Irish translations and associated LS notes, and added to the XML database. Additional multimedia files were added for enhancement within some sections, to illustrate the potential of these file types for the proof of concept outputs.

Three proof of concept outputs were described within Phase IV of work – (1) teacher planning report construction, (2) lexicon creation and (3) handout generation. In each case, the relevant technologies used and the flow of information between the modules were illustrated. Sample screen shots showed the teacher interface in operation.

This chapter has provided evidence of some of the ways that the linguistic content database can be used. It showed how the content can be reused in different proof of concept resource outputs and how the ability to select content *sections* for template-based outputs (which call on different elements of the linguistic content database), can allow the transferability of content to other related contexts. A further enhancement to this process would allow the user direct access to any of the linguistic content database elements.

The next chapter summarises and concludes the research in this thesis, and relates the research journey to the research questions posed in Chapter 1.

Chapter 8 Conclusion

As a visiting primary school teacher of German, the starting point for this research was the more efficient development of language teaching and learning resources, and the empowerment of teachers to break away from the confines of textbooks and other ready-made resources, and provide them with opportunities to more easily tailor and create their own resources for their learners. As a computational linguist, the potential of automatic processing and manipulation of content was evident, and the impetus was to examine how this potential could be transferred to the benefit of these targets. The two domains of education and CL came together to reveal that a beginner's curriculum-bound database of content had features of a controlled language and limited domain, which would allow the easier manipulation of content using CL tools due to the restriction on their language input.

This thesis examined the existing challenges of ready-made resources, and identified the need for more localised resources, tailored to their learners. Regardless of the level of local focus of a ready-made resource, factors such as the processes involved in the publishing industry can skew the original intended content of a resource, and its closeness to a curriculum. Even localised resources will still require some level of localisation by the teacher for their individual groups of learners. Harnessing the potential of technology in the development of content and resources, allows for the reuse of energy- and expertise-expensive content through the strict separation of content from its end resource or product, as well as the enhancement of this content through automatic processing and meta-tagging (e.g. the addition of POS tags). The use of localised settings within these processes, allows for the transferability of relevant content to other related language learning contexts.

It is proposed that the results and processes outlined in this work could be generalised to other contexts in the form of a new development model, called the Tailor Development Model. This model is based on the methods and processes outlined in this thesis and could be transferred for use in other local contexts. The Tailor model could guide the development of language teaching and learning content and its form within a linguistic content database. The model could support the localisation, reuse and transferability of language learning content for specific language learning contexts. As a generalisation from this work, the Tailor Model's use and validation remain to be tested in other research contexts. The Tailor Model will be outlined in Section 8.2 below.

8.1 Research questions revisited

Through the processes of reviewing the research context and the processes of developing a generic database specification and prototype linguistic content database, the following research questions were examined:

1. How can language learning content and resources be localised to particular learning contexts?
2. How can resource content be reused and transferred to other language learning environments?
3. How can a generic linguistic content database be structured?
4. Does linguistic content bounded by a curriculum and localised settings exhibit features of a controlled language?

The following subsections trace the thesis chronology describing the main themes and work around which these research questions are addressed.

8.1.1 The localisation of language learning content and resources

Chapter 2 outlined the need for localised content and resources which are linguistically and culturally relevant to learners. Section 2.5 in particular mapped out the impetuses for localisation and the associated system impact. The localisation industry for digital resources provided an insight into the internalisation process of creating new products which are 'localisation ready'. Chapter 3 noted that the authoring and development processes of consumable resources like textbooks can lack any methodological system and be based more on author intuition and principles, than firm methodologies. Further processes such as the publishing processes can place additional distance between the end product and its intended learner and curriculum focus. CALL development methodologies from Hubbard (1996) and Colpaert (2004) illustrated how localisation can commence through the analysis and design processes of resource generation. The software/translation localisation industry shed further light on processes which can lead to the successful development of localised resources and products, such as the separation of translatable content from its product and the consideration of localisation in the initial design and development-processes.

Research question 1*: How can language learning content and resources be localised to particular learning contexts?*

Chapter 4 outlined the research design and associated methodologies which allowed the development of localised curriculum-based raw linguistic content and associated localised setting notes for the specific research content of the primary school modern language learning classroom in Ireland. Localised settings account for the contextual requirements and parameters of a research context, and were encountered across chapters in this thesis. The research context called for a curriculum-based beginners two year German language learning course which would take account of the other subjects of the national Primary School Curriculum (integration) and links which could be drawn across the modern language curriculum itself (linkage). The local context also called for a more cohesive way to teach languages across the curriculums in place. Learner error analysis was also considered as a relevant factor in localising content and accounting for the presence of three or more languages in the learners' domain. Five localised settings were identified for the primary modern language classroom, which helped to account for the local resource challenges and contextual requirements.

It was the combined processes of devising **curriculum-based** raw linguistic content and accounting for the **localised settings** that localisation was initially accounted for.

The level of success of these development methodologies in accounting for localisation was evaluated through testing the raw linguistic content and associated localised setting notes in a primary school classroom which matched the contextual requirements. The evaluation examined the level of 'learner fit' of the content, which was the degree to which the content matched/suited its intended audience. Few changes to the raw linguistic content were required and the content presented few difficulties to learners. Additional localisation content was also gleaned from the classroom-based research (e.g. typical learner errors) which added to the level of localisation of the content being generated.

The combined processes of **development methodologies accounting for curriculum-based content and localised settings**, in tandem with **classroom-based testing** and addition were successful in accounting for localisation of the content for its context.

A further level of localisation which allows for the reuse of content overlaps research question 2. The internalisation process from the localisation industry proposes that an abstracting or separation of content from its product facilitates the localisation of the language within a product. The separation of content from its product is examined in the next research question.

8.1.2 Ensuring the reusability and transferability of language learning content

Chapter 2 highlighted the pressures of time which teachers experience in juggling the educational requirements of planning, resourcing, teaching and assessing their learners. It further examined how the localisation of digital content can be achieved through the strict separation of content from its product. Chapter 3 similarly outlined how content creation is a time, labour and expertise expensive process. It outlined how few consumable resources maintain content in a format which can be easily extracted for reuse or editing. A finalised paper-based textbook for example, would require the retyping of content in order to reuse it. Its digital cousin would require a similar process of awkwardly copying and pasting formatted content from PDF (or similarly secure formats) into a usable digital format. More technologically-focused development methodologies from the CALL domain and localisation industry highlighted how content can be considered in a separate way to its display and end-product format. All elements which are extraneous to content can be abstracted to alternative development processes, so that content can remain accessible, editable and reusable.

> **Research question 2**: *How can resource content be reused and transferred to other language learning environments?*

Reuse of content

The first important examination of the separation of content from its product was discussed from the outset when the notion of raw linguistic content was introduced in Chapter 1. Its mention in this context was a tentative one, which acknowledged that the separation of language learning content from its surrounding didactics may not be one that sits comfortably among pedagogues and SLA professionals. However, this separation is a temporary one, one which is reintroduced through the language learning teacher being empowered to devise resources from the raw linguistic content and to bring the didactics to the classroom him/herself as the pedagogical professional. The similar concept of content and product separation was also discussed in Section 2.5.2 in the context of the localisation industry, as described above.

Chapter 3 returns to the definition of raw linguistic content and provides a definition in Section 3.1. This definition draws on the work of Decoo (2011), Maley (2011) and Prabhu (1988) who all assert a refocus on content either through beginning with the words of language (Deco, 2011) or by providing more flexible language learning materials by separating content from activities so that they can be recombined in many formats (Maley, 2001; Prabhu, 1988).

The reuse of content is ensured through this separation of content for its reuse in alternative outputs and formats. The use of databases to house content is an established norm. It reflects Colpaert's (2004) three tiered specification system which maintains the separation of the content database, and its administrator and end-user interfaces. It's a similar theme which emerges in other educational research (e.g. Koller, 2007; Ward, 2002) which has a technological focus.

The separation of content from its product was ensured in this research through the development of content in its **raw linguistic form**, as well as the use of a **linguistic content database** to store content. Evidence of the separation is present in the employment of the linguistic content database in Chapter 7 through the three proof of concept outputs for language learning.

Transferability of content

Chapter 4, Section 4.1.3 introduced the concept of localised settings accounting for the transferability of content from one language learning context to another. In accounting for the contextual requirements and parameters of a particular context, they can also account for parameters which are applicable to other contexts. Through switching the localised settings on or off, relevant content which is reusable in related contexts can be accessed and employed. The localised settings similarly can be reemployed in an indirect way in a scenario similar to that used in the internalisation process from the localisation industry, where relevant content for language learners can be translated into alternative target languages. Chapter 7 highlighted how the localised settings could be accessed and either used or discounted by the teacher creating their own resources. Through the use of localised settings, the transferability of content was ensured.

The alignment of a curriculum to an international standard or benchmark can also allow for the reuse of a certain amount of language learning content. The modern language curriculum documents in this research context were aligned to level A2 of the Common European Framework of Reference for Languages (CEFRL), which also makes the content transferrable to similar language learning contexts.

8.1.3 Specifying a generic linguistic content database

The use of a database for storing content is a useful tool in maintaining the separation of content from its product. There are many types of databases, where the structure and format of their labelling schemes differ, as well as the means used to access the content. The research design and methodologies outlined in Chapter 4 described the processes which led to the specification of the linguistic content database. The classroom-based research concluded these processes and the specification for the linguistic content database was outlined.

Research question 3: *How can a generic linguistic content database be structured?*

Chapter 7 outlined the rationale for the selection of XML as the mark-up language chosen to describe the content of the linguistic content database. The structure used provides scope for alternative localised settings to be added, and allows the integration of alternative/additional languages.

8.1.4 Controlled languages and limited domains in language learning content

The parallel thread which wove through the thesis discussions was an examination of how raw linguistic content could be analysed, characterised and compared to alternative types of language learning resources. Chapter 6 outlined the matrices from SLA and CL which were employed for these purposes. The SLA matrices are usually employed by teachers or practitioners to assess progress made by their learners. In this instance, they were employed to investigate features of controlled languages across the three datasets as well as to describe the language contained within them.

Research question 4: Does linguistic content bounded by a curriculum and localised settings exhibit features of a controlled language?

The initial focus of the analysis was on proving that the raw linguistic content which had been created differed from other types of language resources, hence making the development process worthwhile, as a new 'type' of content could be created. The result of the development process was a localised discrete entity, specific to its contextual requirements and parameters. While not directly related to the development of localised, reusable or transferrable language learning content, the secondary focus of the analysis lay in examining whether the raw linguistic content exhibited any controlled language features. The further reaching implications of the raw linguistic content having controlled language features lie in the easier processing and manipulation of the content using CL/NLP tools as the language input to these systems is restricted. This could lead to greater success for ICALL systems which rely on the robustness of the integrated CL/NLP technologies for their own success.

Chapter 6 outlined how the raw linguistic content does exhibit features of a controlled language. Only one CL tool was used within this research – the TreeTagger POS tagger (Schmid, 1994). It performed with a tagging accuracy of 95%, which fell a little short of the 96.36% tagging accuracy reported by Schmid for the tagger (1994). The shortfall in accuracy was somewhat

accounted for by the embedding of English language in section headers and optional markers in sentences which could not be accounted for within the German lexicon and corpus the POS tagger is trained on.

8.2 Generalising research results and processes to other contexts: the Tailor Development Model

The above results can be formalised as a new development model, which generalise the methods and processes outlined in this research and allow them to be transferred for use in other local contexts. The new model accounts for the development of localised, reusable and transferrable teaching and learning linguistic content databases. It is a formalisation of the development processes and methodologies which were outlined in Chapter 4. As such, existing development principles, frameworks and models inform the development process where relevant and result in the emergence of a new model to support the development of linguistic content databases. No one existing methodology or framework examined through this research provides a full account of the development processes which were required to develop such a system, hence the result in formalising the processes outlined.

The new development model is named the Tailor Development Model (TDM), and can be used to develop teaching and learning content for any local setting or context. *Tailor* is a quasi-acronym, with the letter 'T' representing 'Teaching', 'L' representing 'Learning' and 'R' representing 'Resource'. As a word in its own right, *Tailor* represents a person who measures, and then cuts, and joins disparate elements together to form a new whole. Tailors have also traditionally been able to take one bolt of fabric, and make many types of garments as outputs.

The new model differs slightly to the processes outlined in Chapter 4, in that Stage 2 of the Tailor Development Model does not represent an analysis and characterisation stage, rather a stage to process and enhance the content, potentially using CL/NLP tools before it is formalised into a database in Stage 3. Similarly, Phase IV of the thesis research design does not carry through to the Tailor Development Model as a proof-of-concept stage.

8.2.1 Proposal for a new development model

It is proposed that the Tailor development model (TDM) be is divided into three stages of work (see Table 8.1). Each stage reflects a transition in the format of the raw linguistic content, usually as a result of a process, or series of processes in any one stage which impact the raw linguistic content and the LSs. Stage 1 sees the initial development and evaluation of the raw linguistic content and LSs to determine their fit to the local context. Stage 2 sees the alteration and/or

addition of content through processing and enhancement of the raw linguistic content and LSs (addition of POS-tags for instance). Stage 3 sees the raw linguistic content and associated LSs being transformed from a listing of content and associated data to a searchable and didactically usable database.

TDM Stage of work	Description
Stage 1	Development of raw linguistic content and evaluation in the local context
Stage 2	Processing and enhancing raw linguistic content
Stage 3	Compilation of all development components and development of the reusable linguistic content database

Table 8.1 Stages of work comprising the Tailor Development Model

The stages of work from the research design and the Tailor Development Model (TDM) roughly align to one another (see Figure 8.1). Phase I of the research design aligns to Stage 1 of the TDM, aside from Process ‘d’ which becomes part of Stage 3 of the TDM. Processes ‘b’ and ‘c’ of the research design are accounted for in the TDM through the feasibility considerations (see Stage 1 below), the local research context and the research methods employed. These processes will be specified by the local context rather than be specifiable within the TDM.

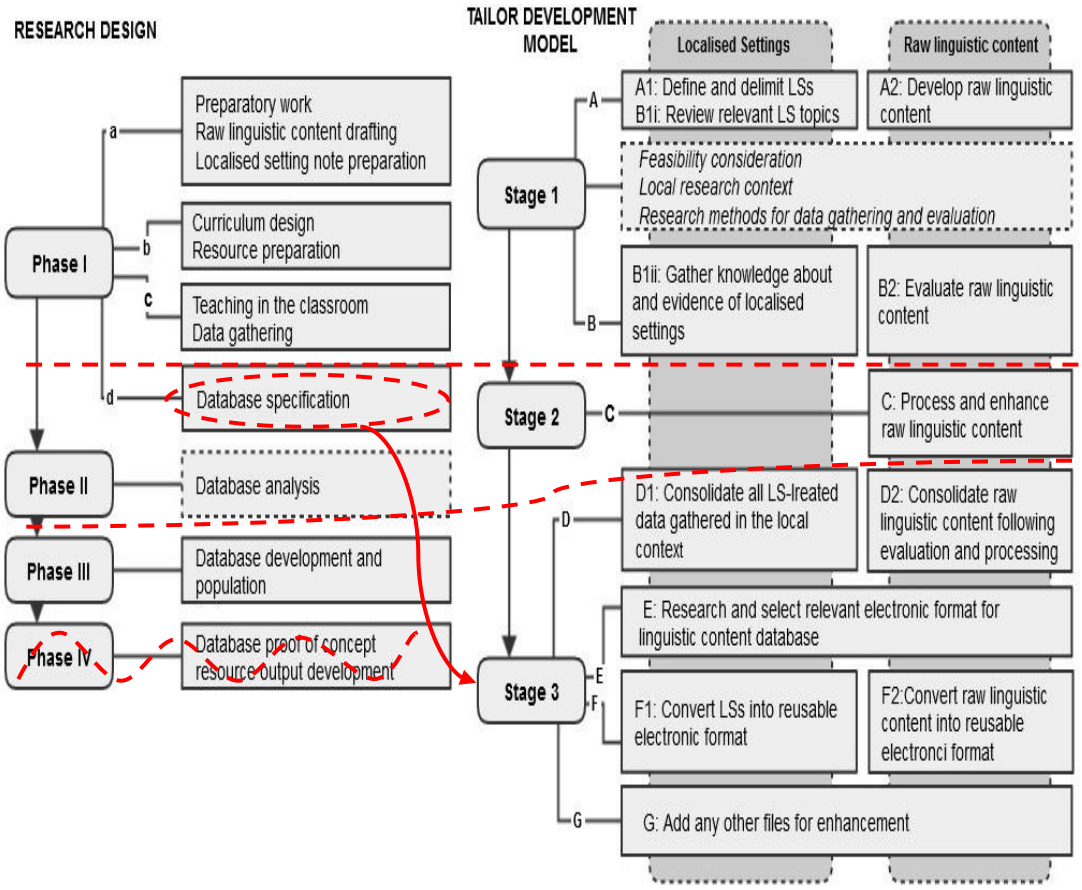


Figure 8.1 Alignment of the research design with the Tailor Development Model

As outlined above, Phase II of the research design aligns to the TDM, but it is only the processing stage that is relevant to the TDM rather than the analysis stage. Phase III of the research design aligns to Stage 3 of the TDM, as the data is consolidated, and the database is developed and populated. Phase IV of the research design which involves the proof of concept outputs of the linguistic content database is not relevant to the TDM.

By moving or removing elements of the research design, a new model can be proposed which attempts to generalise the research processes and methods to other contexts. The relatively direct transfer of the research design phases to TDM stages allows the relevant findings of this research to also be relevant to the TDM. The Tailor Development Model can guide the development of localised, reusable and transferrable linguistic content databases for specific local contexts.

8.2.2 Tailor: Stage 1

It is proposed that the first stage of work in the development model be divided into two broad tasks, A and B:

A. Development work prior to entry into the local context
    1. define the local context and the boundaries of the localised settings as per the local context
    2. develop the initial raw linguistic content
B. Evaluate the raw linguistic resources within the local context
    1. gather the required information for the localised settings which are defined by the local context
    2. evaluate and revise the raw linguistic content in the local context.

Task A takes place outside of the classroom or local context. The raw linguistic content is drafted in preparation for evaluation through Task B in the local context. Similarly, the localised settings are defined and delimited during Task A, before entry into the local context. This ensures that all factors which need consideration have been mapped out before the evaluation in the local context commences to fulfil Task B (see Figure 8.2, p. 251).

Stage 1, Tasks A1 and A2

The goal of Task A1 is to pin down the factors which are to be included in the considerations of the content system to be developed.

The ***local context*** is the language learning environment for which the reusable resource system is to be created. The definition of the local context is what identifies any one development model iteration output from another – who the resources created are suitable for (and transferrable to).

Examples of the elements the definition could include are:

- the target learner (who?)
- the language level/competency to be attained through using the resource
- the educational setting (e.g. early childhood, primary, post-primary, secondary or tertiary education)
- any formal national curriculum or international benchmark which the language learning materials will/can be aligned to
- the language(s) of focus/to be taught (e.g. French, Italian).

The local context is further defined through the ***localised settings*** **(LSs)**. Following the same use of the term for the research methodology, localised settings vary from one local context to another and are those elements which are investigated prior to entering the local context and about which information is gathered during Stage 1 of the development model (factors relating to the '*who'*, and the '*what'*). LSs are used to localise the raw linguistic content and content database to any local teaching and learning environment and to delimit the extent of the evaluation to be carried out in the local context setting. These are the variables or factors which will be examined in the local context.

Examples of localised settings could include details of

- the target audience/learner profile (e.g. age, gender, linguistic background) – these can be real or personas, depending on the developer – if the developer is a practitioner then these would be real as he/she would be developing a solution to his/her own teaching need. If the developer is a publishing house, then the initial definition could be based on personas, which are realised as real people in a classroom when it comes to defining the factors connecting Tasks A and B (the research space participants).
- the language area to be taught (this could be as narrow as 'hobbies' or as broad as 'Junior Certificate content for Gaeilge')
- the typical language learning errors of your target audience
- the end content system users (which may dictate some of the factors connecting tasks A and B, e.g. research duration).

Further examples are summarised in Table 8.2.

Sample Localised Settings	Examples
Target audience/learner	Age, gender, education level, environmental/home/school factors, native language, other known language(s), prior knowledge/learning, Cultural norms, socioeconomic factors (e.g. designated disadvantaged)
Language area to be taught	Introductions, Engineering language, the GCSE Gaeilge syllabus, to level B1 on the Common European Framework of Reference for Language (CEFRL)
Typical language learning errors	False friends, cognates, inflectional errors, plural errors
End content system users	Teachers, researchers, practitioners, learners, publishing houses, academics, software engineers, etc.

Table 8.2 Sample LSs and their associated examples

It would be impossible to examine every variable facet in any one local context which could impact on teaching and learning. Therefore, defining and delimiting LSs *before* entry into the local context ensures that (1) the local context research of Task B is focused on only these variables, and (2) efforts do not stray to additional variables for which there is no capacity in the time available in the local context setting to examine. Defining and delimiting these variables also provides some gauge of the extent of time which will be required in the local context. Such time can often be hard for researchers to acquire, especially in educational settings which are already hard pressed for time (e.g. NCCA, 2005a, 2008a). If a suitable length of time cannot be acquired in the local context, the localised settings (LSs) can be trimmed until their requirements on time suit what is available in the local context. Colpaert (2004) defined "feasibility considerations" (p. 112) as those considerations such as funding and man-power which may involve lowering or raising aspirations.

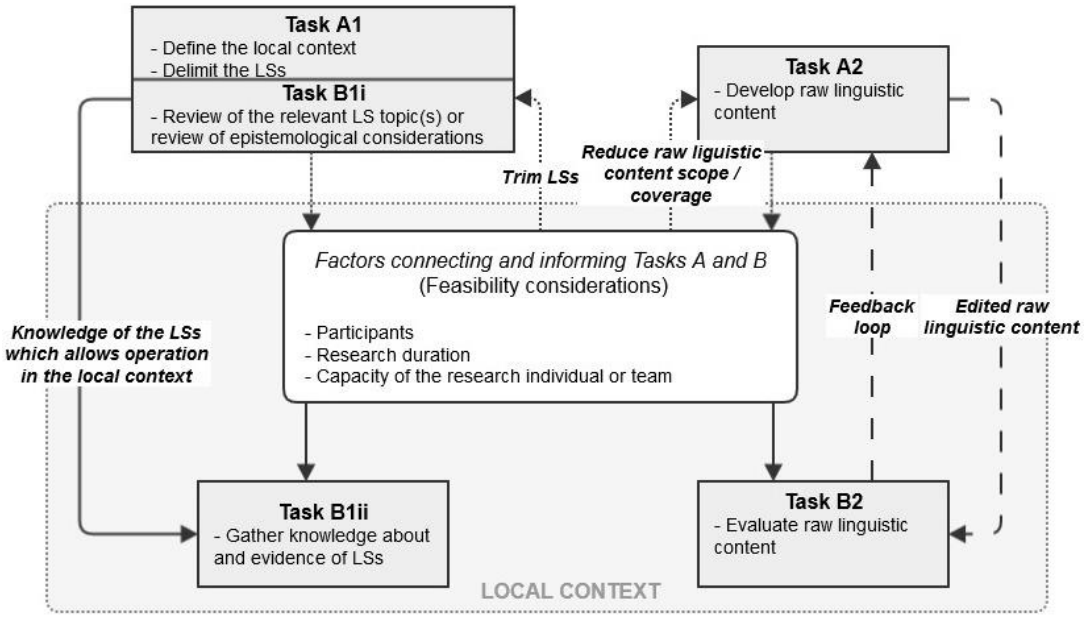


Figure 8.2 Stage 1 of the Tailor Development Model: Interconnectedness of the tasks

Stage 1 Task A2 of the Tailor Development Model involves **developing the raw linguistic content**. This forms the base or core content words, phrases and sentences from which the full content database will later be developed. It comprises the words, phrases and sentences which could emerge through teaching to the local curriculum. This raw linguistic content is evaluated in the learning environment specified in the local context. Raw linguistic content is transferrable, in that it can be used in any language learning environment where the same language is being learned to the same level. Taking care in their compilation and aligning them to international benchmarks (e.g. the CEFRL) will ensure a greater level of transferability to other language learning environments. Naturally, a big difference in the profile of the learner may have an impact on the suitability of the materials. If the learners are children, the register of the resources may need to be adapted, as will some of the content, e.g. how to get to school (Child: bus, walk, cycle, driven) vs. how to get to work (Adult: bus, walk, cycle, drive). However, modelling on an international benchmark will provide the necessary baseline language (which can be added to to cater for alternative learner profiles) to reach the level of proficiency specified. Table 8.3 outlines the principles which underpin the development of raw linguistic content. The compilation of principles 1-3 in this table emerge from the need to address the resourcing challenges identified in this research. Principles 4 and 5 proved useful stages of work in the course of this research.

The raw linguistic content could:
1. be based on the relevant national/global curriculum, where relevant
2. support any national/global language programme in place, if any
3. be relevant to the students they are designed to cater for, as far as possible
4. be proofed by another teacher/researcher/native speaker before being used in the classroom
5. be evaluated in a classroom setting before being formalised (Task B2).

Table 8.3 Principles underpinning the development of raw linguistic content

The development of raw linguistic content depends on the expertise of the individual or among the development group, the resources available to them and the time available.

Sources which could be consulted and employed to develop the raw linguistic content include:

- the local curriculum or standard benchmark (if part of or relevant to the local context)
- any relevant content from supporting texts in the local context, or in similar global environments
- electronic sources like dictionaries, thesauri, websites or corpora.

The level of complexity for this content will depend on the language level of the learners, curriculum or benchmark.

Factors connecting and informing Tasks A and B: Participants, research duration and research design

The details of the ***participants, research duration and research design*** act as connecting factors between Tasks A and B (see Figure 8.2). Usually, these factors are determined at the outset of any research or study. In the instances of the Tailor Development Model, they may be known at the outset of development work if the developer is also a practitioner who intends testing with his/her own learners. However, if the developer is a publishing house, for instance, these factors may not be determined until a suitable target research space is identified in which trialling can take place. In both instances, these factors often influence elements of Task A and the extent and focus of the local context-based research of Task B.

Stage 1: Tasks B1 and B2

With the basis of the Tailor Development Model in Design-Based Research, a mixed methods approach can be taken to conducting Task B – gathering knowledge about and evidence of the LSs and evaluating the raw linguistic content. Task B1 involves **gathering knowledge about and evidence of localised settings**. Task B1 is further subdivided into two broad actions – (1) review of the relevant LS topic or review of epistemological considerations and (2) evaluation in the local context. Action (1) is best carried out before entry into the local context (as mentioned above) so that knowledge learned about the LSs can be applied or incorporated into the local context-based research. An example of these processes at play would be (a) researching learner errors and error analysis (b) incorporating an error tagging schema which was informed from researching the field.

Task B2 involves the **evaluation of the raw linguistic content in the local context**. This is an iterative cycle, which leads back to task A2 and revisions to the raw linguistic content (where time in the local context allows). The raw linguistic content can be tested for completeness (is the content exhaustive enough for learners in the local context?), accuracy (is the content grammatical and comprehensible?), applicability (is the content appropriate to the learners?) to the local context and suitability for purpose. Research methodologies which can be employed in the evaluation of the raw linguistic content include observation and data gathering (learners' production). This iterative process presents a formative evaluation process for the raw linguistic content (see feedback loop in Figure 8.2).

8.2.3 Tailor: Stage 2

After completing Stage 1, all local context research work would be complete and the next stage of work would involve processing and enhancing the content to maximise its potential when in use.

The degree and type of processing in Task C depends on the time available, the level of requirement for adding annotation to the raw linguistic content and the target use or output of the linguistic content database under construction. It could also depend on the level of expertise of the practitioner or among the research team. Processing which could be carried out includes word frequencies or tokenisation. Additional annotations such as morphological annotations, POS tags, translations or synonyms could also be added. Processing at this stage need not involve the use of CL/NLP tools. It may simply be a reorganisation of content following time in the local context or a transfer of content into an alternative format.

The next section outlines Stage 3 of the Tailor Development Model, where all of the developed and processed components are compiled into the final linguistic content database.

8.2.4 Tailor: Stage 3

Stage 3 of the Tailor Development Model draws all of the content which can be authored, tested and processed across Stage 1 and Stage 2 into the usable linguistic content database (Figure 8.3). The priority of this stage of work is to maximise the output from the Tailor Development Model to ensure it is fit for the purpose for the local context it was designed to cater for.

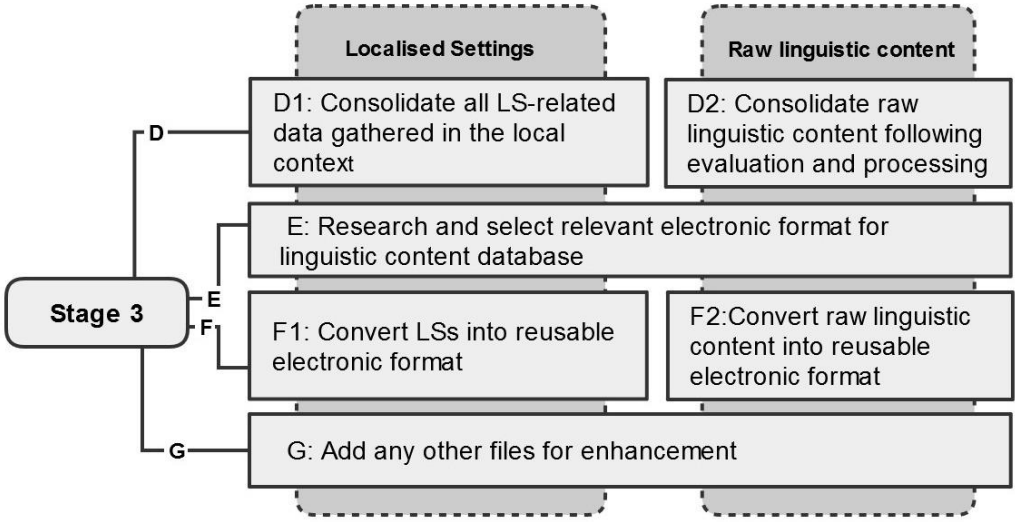


Figure 8.3 Stage 3 of the Tailor Development Model

Task D1 involves consolidating all of the data gathered on the LSs in the local context. While this data will have been recorded as part of the research in the local context, it may be necessary to take stock of the data that has been gathered and decide which, if not all, of the data will be incorporated or taken into account in the final linguistic content database. Similarly, the raw linguistic content may have been altered through additions from its evaluation in Stage 1 and/or processing in Stage 2. **Task D2** involves examining the raw linguistic content to ascertain which raw linguistic content is relevant for inclusion in the linguistic content database. The output of these tasks is a listing of all of the datatypes which comprise the raw linguistic resources and associated LSs and how they relate to the linguistic content database.

The format of the content database is an important decision to make through **Task E1**. Depending on the end use of the content database and what potential consumable resources it will be used to create, the format chosen will need to be compatible with the output system which will call on it. If a large amount of data constitutes the raw linguistic content and associated LSs, then the performance of the database type will need to be considered (to search for and extract relevant content in an acceptable amount of time).

Task	**Process/Function**	**Guiding questions** for considering the best technological format and content for the content database
E	Selection of relevant electronic format/ mark-up	• What type of format is relevant to the use of the content database? • Which format type will be most accessible and usable to the end user? • Are there any existing output systems which need to be catered to? • How large is the data? (relevant to the performance of the database) • Will any flexibility be required to add/delete content?
F1	Addition of fields for LSs	• Are all of the LSs complete and usable? (Were they all fully researched in the local context?) • How can the LSs best attach to their raw linguistic content? What raw linguistic content are they most relevant to?
F2	Conversion of raw linguistic content into a reusable electronic format	• Are there any technological restrictions in the design space? • Are there any limits in technological skill-set among the research individual or team? • How will the content be transferred to the database? (How will the database be populated?) • How long will this process take?
G	Addition of any other files for enhancement (e.g. images, audio)	• Are there any multimedia files (audio, image, video) available which could enhance the database?

Table 8.4 Guiding questions to scaffold Stage 3 of the Tailor Development Model

Task F2 generally precedes **F1**, as the raw linguistic content format will influence the layout of the database and the LSs are subsequently added to the raw linguistic content in the database. Transferring content into a database system can be a time consuming process, especially if there are large amounts of data. Interfaces are available to ease this process for most database or mark-up languages.

The final task, **Task G**, provides for the consideration and addition of any multimedia files which are relevant to the content database and will enhance the final content system, e.g. associated images, audio files or videos. Table 8.4 provides some guiding questions to scaffold the work in this final phase.

Once the linguistic content database has been compiled, further considerations will include how the end user/practitioner will access the data, how the data will feed into the system outputs (e.g. a template-based handout (paper-based or electronic), a teacher report for planning or a topic-based lexicon) (see Chapter 7).

8.3 Limitations of the research

The individual nature of this research journey

Carrying out the research in this thesis as an individual, rather than a member of a research team had some limiting factors. The processes of teaching and gathering data at the same time as described in Chapter 4 for Phase I of work limited the extent to which data could be gathered and recorded as it was observed. The one hour delay in noting observations may have led to fewer observations being made than if the observations could have been noted immediately.

Similarly, the processes of evaluating the raw linguistic content and associated localised settings in Chapter 5 are centred on an inverse process from their design and development (Hubbard, 1996). As described, their learner fit is determined by checking that the content is fit for purpose in its intended context. The data gathered was somewhat circular in that the results of using it in the classroom were very positive, with few additions noted for the raw linguistic content. As I was the content creator and evaluator, it could be argued that I was biased towards the relevance of the content being evaluated. A more neutral evaluator may have shed different light on the relevance of the raw linguistic resources. On the other hand, considering that the raw linguistic content and associated localised setting notes were developed for the particular language learning context they were evaluated in, it could also be the case that the initial development processes were suitably targeted to warrant such low levels of adjustment being evidenced through the evaluation.

Design quality of proof of concept outputs

Harbison (2008) described how teachers' own resources cannot compete with professionally-designed textbooks for interest, colour or appeal. The sample proof of concept outputs of Chapter 7 are far removed from the vibrant designed textbooks which are in use in classrooms. The purpose of the proof of concept outputs was to show the potential resource outputs which could be created from the one content source. While their content is very relevant to the research context of the primary modern language learning classroom in Ireland, their visual appeal is very low. Designers and computer programmers are better placed to integrate the content into more attractive end product resources through pure design and/or programming of additional output formats. One suggested area for future work would be to design a multitude of output templates or resources which could stand side by side with professionally-designed materials.

Database content for the English and Irish translations of the raw linguistic content

The raw linguistic content was developed in the target language, German. English and Irish translations were incorporated to assist in identifying areas of Language Awareness and to further enhance the content. The English and Irish translations were not tagged with POS tags, and therefore do not align directly to their German equivalents. However, the structuring of the database into word, phrase and sentence entities does allow the relevant information to be extracted in all languages where certain outputs are under development, such as a lexicon. The database structure also allows for the addition of POS tags in Irish, English or any other relevant languages. An enhancement to the current content would be the addition of POS tags for English and Irish to allow for greater processing and alignment of languages.

Validating the new Tailor Development Model

The new Tailor Development Model was realised as a generalisation of the results and processes outlined in this thesis. As a result, it has not been evaluated as an entity in itself. The use and adoption of the Tailor Model would need to be preceded by a validation of the model through its testing in other contexts.

Accounting for sequencing in larger linguistic content databases

The proof of concept linguistic content database was relatively small in its content makeup. As such, the use of linkage notes to indicate a broader reference to sequencing for learning (i.e. which preceding and future sections the present section was related to) sufficed as indicators of sequencing. Larger databases would be more difficult to reference in this way. Therefore, a more specific system would be useful for assisting teachers with sequencing across the database content.

8.4 Suggestions for further research

The methods and processes outlined in this work were generalised to form the new development model, Tailor. As such its use and validation in other contexts remains to be tested. This testing in an alternative context could provide an avenue for further research.

The focus of this research was on the development of localised, reusable and transferrable language learning content which could be employed in the creation of multiple language learning outputs or resources. Three proof of concept outputs were illustrated in Chapter 7. One area of further research would be the examination of additional outputs to employ the developed content, potentially with a greater emphasis on design (as outlined above).

A further employment of the developed linguistic content database would be an examination of the success rates of other CL/NLP tools on the raw linguistic content, given that it exhibits features of a controlled language. The results could determine suitable CL/NLP tools which could be used in tandem with the content to develop ICALL-type language learning resources and tools.

The new integrated primary language curriculum for English and Irish under development in Ireland (NCCA, 2014) includes a learning continuum which describes how children progress in their language learning skills across eight levels labelled ‘a’ to ‘h’. This approach is more in keeping with the descriptors used in the CEFRL. The SLA matrices used and described in Chapter 6 could be employed to profile the language gains made from one section of the linguistic content database to another, and place the learning on the new curriculum continuum. This would have the dual benefit of aligning the modern language curriculum document content to tie in with the new integrated one for English and Irish, and placing the English and Irish content within the linguistic content database on the learning continuum to illustrate progression across the sections

This thesis outlined how there are many children for whom English is an additional (EAL) attending schools in Ireland. As well as the linguistic wealth and benefit these children and their families introduce to school environments, they also present new languages through which the school and its staff may not be able to communicate. School budgets do not stretch to professionally translating home correspondence to these children’s home languages. Often, schools rely on free online MT systems to correspond with families in languages other than English and Irish. Where these systems are good for providing the gist of a text, their use in the school instance could lead to confusion through inaccurate translation of content. The concept and processes of developing the linguistic content database outlined in this work, could be

transferred to the development of a linguistic content database which accounted for the frequent types of correspondence which are sent to school children's homes. There is a relatively finite listing of these occurrences, such as school closures, meetings, health alerts, reports and notifications. Compiling the language required for these areas of correspondence and having them centrally translated into the most common of the new languages, would provide a database of usable content which could be accessed and used to create correspondence for home.

8.5 Moving forward

As a concluding reflection, a very recent meeting with primary school teachers, teacher educators and national curriculum advisors and creators revealed that the same identified issues remain pertinent in educational settings. Some teachers in Ireland are overloaded with the sheer depth and breadth of curriculum documents, planning requirements and textbook-heavy classrooms. Other teachers seem to cope better with the document-heavy requirements. Some long-standing teachers who have seen much curriculum change and reform in the past 15 years, in comparison to the uneventful years between 1971 and 1999 when the curriculum remained unchanged, are eager for a less tremulous period with fewer changes.

Providing a clear path for teachers to navigate through the myriad of requirements from curriculum, to planning, to resourcing and teaching in the classroom, is a priority for educational stakeholders. Easing the burden somewhat for over-stretched teachers through the easier access to and provision of relevant curriculum-based resources, would be a small stepping stone along the long path of educational requirements within teachers' remit.

References

Anderson, T. and Shattuck, J. (2012). Design-Based Research: A Decade of Progress in Education Research? *Educational Researcher 41*(1), 16-25.

Amaral, L. A. and Meurers, D. (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL 23(1),* pp. 4-24.

Apple, M. W., (1985). The Culture and Commerce of the Textbook. *Journal of Curriculum Studies, 17(2),* pp. 147-162.

Arnold, D.J., Balkan, L., Meijer, S., Lee Humphreys, R., and Sadler, L. (1994). *Machine Translation: An Introductory Guide.* London: Blackwells-NCC.

Arnold, W. and Rixon, S. (2008). Materials for Teaching English to Young Learners. In B. Tomlinson, (Ed.), *English Language Learning Materials: A Critical Review*. London, New York: Continuum International Publishing Group, pp. 38-58.

Axbey, S. (1989). *Standard exercises in self access learning*. Presentation at British Council Course on Self-Access Learning, Cambridge, May 1989.

Ball, D. L. and Feiman-Nemser, S. (1986). Using Textbooks and Teachers' Guides: What beginning elementary teachers learn and what they need to know. Research Series No. 174. Michigan State University: Institute for Research on Teaching.

Barnardos (2012). *School Costs Survey 2012*. Ireland: Barnardos. Available online at: http://www.barnardos.ie/assets/files/Advocacy/2012-School-Costs-Report.pdf [Accessed: 17/11/12].

Bacha, N., Ghosn, I. and McBeath, N. (2008) Materials used in the Middle East. In B. Tomlinson (Ed.) *English Language Learning Materials: A Critical Review*. London, New York: Continuum. pp. 281-299.

Barrios, M. L., de Debat, E. V. and Tavella, G. (2008). Materials in use in Argentina and the Southern Cone. In B. Tomlinson (Ed.) English Language Learning Materials: A Critical Review. London, New York: Continuum. pp. 300-316.

Bell, J. and Gower, R. (2011). Writing course materials for the world: A great compromise. In B. Tomlinson (Ed.) *Materials Development in Language Teaching (2nd Ed.)*. Cambridge: Cambridge University Press, pp. 135-150.

Blin, F., Jalkanen, J., Taalas, P. and Nic Giollamhichil, M. (2013). Towards a sustainability framework for large collaborative CALL projects. In *Conference proceedings from WorldCall 2013: Global perspectives on Computer-Assisted Language Learning.* Glasgow, UK: University of Ulster, pp. 29-31.

Bolitho, R. (2008). Materials for Central and Eastern Europe. In B. Tomlinson (Ed.) *English Language Learning Materials: A Critical Review*. London, New York: Continuum. pp. 213-222.

Bolitho, R., Carter, R., Hughes, R., Ivanic, R., Masuhara, H. & Tomlinson, B. (2003). Ten questions about language awareness. *ELT Journal* 57(3), pp. 251-259. UK: Oxford University Press.

Bongers, H. (1947). *The History and principles of vocabulary control, as it affects the teaching of foreign languages in general and of English in particular* (3 vols.). Woerden: Wocopi.

Brants, S., Dipper, S., Eisenberg, P., Hansen-Schirrai, S. König, E., Lezius, W., Rohrer, C., Smith, G. and Uszkoreiti, H. (2004). TIGER: Linguistic Interpretation of a German Corpus. *Research on Language and Computation, 2*, pp. 597-619.

Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL 17(1*). pp. 47-64. UK: Cambridge University Press.

Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven learning. *ReCALL 19(3*). pp. 307-328. UK: Cambridge University Press.

Breen, M. P., Candlin, C. N., and Waters, A. (1979). Communicative materials design: Some basic principles. *RELC Journal, 10*, pp. 1-13.

Brezina and Gablasova (2003). Is There a Core General Vocabulary? Introducing the New General Service List. *Applied Linguistics 2013*, p. 1-23. Available online at: http://applij.oxfordjournals.org/content/early/2013/08/25/applin.amt018.full.pdf+html [Accessed 12/07/'14].

Brown, H. D. (1994). *Principles of Language Learning and Teaching.* London, Sydney, Toronto, Mexico, New Delhi: Prentice-Hall Inc.

Browne, C., Culligan, B. and Phillips, J. (2003). A New General Service List. Available and maintained online: http://www.newgeneralservicelist.org/

Brusilovsky, P., Knapp, J. and Gamper, J. (2006) Supporting teachers as content authors in intelligent educational systems, *Int. J. Knowledge and Learning*, Vol. 2, Nos. 3/4, pp.191–215.

Burke, M. (2002). *Optimised Algorithms and Data Structures for Spelling Error Detection and Correction.* MSc Thesis, Trinity College Dublin, Ireland.

Butler, D., Leahy, M., Shiel, G. and Cosgrove, J. (2013). *Building towards a learning society: A national digital strategy for schools. A consultative document*. Dublin: Department of Education and Skills.

Byrnes, H. (2001). Articulating foreign language programs: The need for new, curricular bases. In C. G. Lally (Ed.), *Foreign language program articulation: Current practice and future prospects.* Westport, CT: Bergin and Garvey, pp. 157-180.

Cadwell, P. (2008). Readability: Examining its usefulness in the field of controlled language. *Localisation Focus 11(1).* pp. 34-45.

Callaway, D., (1985). *Washington State Adult Refugee ESL Master Plan* (Revised). Seattle, WA: Department of Public Instruction/Department of Social and Health Services.

Candelier, M. (Ed.) (2004). *Janua Linguarum. The gateway to languages. The introduction of language awareness into the curriculum: awakening to languages*. Strasbourg Cedex: European Centre for Modern Languages, Council of Europe Publishing.

Child, D. (1986). *Psychology and the Teacher.* 4th Edition. London, New York, Sydney, Toronto: Holt, Rinehart and Winston.

Clark, J. I., (1987). *Curriculum Renewal in School Foreign Language Learning*. Oxford: Oxford University Press.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R. & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher, 32*(1), 9-13.

Collins, G. (2012). *How Texas inflicts bad textbooks on us*. The New York Review of Books. June 21 2012. Available online at http://www.nybooks.com/articles/archives/2012/jun/21/how-texas-inflicts-bad-textbooks-on-us/. [Accessed: 19/01/15]

Cook, V. (2012). *Second Language Learning and Language Teaching* (4th Edition). London: Hodder Education.

Cooper, A. (1997). *The inmates are running the asylum: Why high-tech products drive us crazy and how to restore the sanity*. Indianapolis, IN: SAMS.

Colpaert, J. (2004). Design of Online Interactive Language Courseware; Conceptualization, Specification and Prototyping. Research into the impact of linguistic-didactic functionality on software architecture. Published Ph.D. thesis, University of Antwerp, Belgium.

Corder, S. P. (1974). Error Analysis. In J. P. B. Allen & P. Corder (Eds.), *The Edinburge Course in Applied Linguistics. Volume 3 – Techniques in Applied Linguistics* (pp. 122-131). London: Oxford University Press.

Council of Europe and Department of Education and Science (2008). *Language Education Policy Profile: Ireland*. Available online at: http://www.coe.int/T/DG4/Linguistic/Profils1_EN.asp [Accessed: 20/03/08]

Council of Europe. (2001). Common European Framework of reference for languages: Learning, teaching, assessment. Cambridge: Cambridge University Press.

Council of Europe. (2006). *European Language Portfolio: Key reference documents*. Available online at: http://www.coe.int/T/DG4/Portfolio/documents/keyrefdocs.doc [Accessed: 20/03/08].

Creswell, J. W. & Plano Clark, V. L. (2007). *Designing and Conducting Mixed Methods Research*. California, UK, New Delhi, Singapore: SAGE publications.

CSO (Central Statistics Office) (2011). Census 2011 Profile 6 Migration and Diversity - A profile of diversity in Ireland. Available online at: http://www.cso.ie/en/census/census2011reports/census2011profile6migrationanddiversity-aprofileofdiversityinireland/ [Accessed: 05/11/14].

Curtis, H. (2005). *Passwort Deutsch*. Stuttgart: Ernst Klett Sprachen GmbH.

Dat, B. (2008). ELT Materials Used in Southeast Asia. In B. Tomlinson (Ed.) *English Language Learning Materials: A Critical Review*. London, New York: Continuum. pp. 263-280.

Davies G. (2011) CALL Software and Website Evaluation Forms. In G. Davies (Ed.) *Information and Communications Technology for Language Teachers (ICT4LT)*, Slough, Thames Valley University [Online]. Available at: http://www.ict4lt.org [Accessed 10/01/15].

De Posada, J. M. (1999). The presentation of metallic bonding in high school science textbooks during three decades: Science education reforms and substantive changes of tendencies. *Science Education, 83*(4), pp. 423-447.

Decoo, W. (1996). Didascalia: Background and Basic Options. *Computer Assisted Language Learning, 9* (4). p. 299-308.

Decoo, W. and Colpaert, J. (1996). Guest Editorial: Engineering in Language Teaching and Testing: The Experience of the Didascalia Research Centre. *Computer Assisted Language Learning, 9* (4). p. 297-298.

Decoo, W. and Colpaert, J. (1997). The Dynamic Relation between Textbook and Software in a Research Based Language Learning Method. In K. A. Murphy-Judy (Ed.) *NEXUS The Convergence of Language Teaching and Research Using Technology.* CALICO Monograph Series Volume 4. pp. 66-77. North Carolina: CALICO.

Decoo, W., Heughebaert, E., Schonenberg, N. and van Elsen, E. (1996). The Standard Vocabulary Programs of Didascalia: In Search of External Versatility and Didactic Optimisation. *Computer Assisted Language Learning, 9*(4). pp. 319-330.

Decoo, W. (2011). Systemization in Foreign Language Teaching: Monitoring Content Progression. New York, Oxon: Routledge.

Delcloque, P. (1998). Chapter 1: Design Methodologies. In F. Blin, N. Chénik & J. Thompson (Eds.), *CALL Courseware Development: a Handbook.* University of Hull: EUROCALL.

Department of Communications, Energy and Natural Resources (DCENR) (2013). *Doing more with digital. The national digital strategy for Ireland. Phase 1 Digital Engagement*. Dublin: The Department of Communications, Energy and Natural Resources.

Department of Education and Science (1998). *Schools IT2000 - A Policy Framework for the New Millennium*. Dublin: The Government Stationery Office.

Department of Education and Science (1999). Walk Tall: Social, Personal and Health Education: For the Prevention of Substance Misuse. Dublin: The Government Stationery Office.

Department of Education and Science (2000). Blueprint for the Future of ICT in Irish Education. Three Year Strategic Action Plan 2001 – 2003.

Department of Education and Science, (2005). An Evaluation of Curriculum Implementation in Primary Schools: English, Mathematics and Visual Arts. Dublin: The Government Stationery Office.

Department of Education and Science, (2005a). Literacy and Numeracy in Disadvantaged Schools: Challenges for Teachers and Learners. An Evaluation by the Inspectorate of the Department of Education and Science. Dublin: The Government Stationery Office.

Department of Education and Science, (2005b). *DEIS* (*Delivering Equality Of Opportunity In Schools). An Action Plan for Educational Inclusion*

Department of Education and Science, (2007). *Irish in the Primary School. Inspectorate Evaluation Studies.* Dublin: The Government Stationery Office.

Department of Education and Science, (2008). ICT in Schools. *Inspectorate Evaluation Studies.* Dublin: The Government Stationery Office.

Department of Education and Science (2008a). *Investing Effectively in Information and Communications Technology in Schools 2008-2013*. Ireland: DES.

Department of Education and Science, (2009). *Social, Personal and Health Education (SPHE) in Primary Schools Inspectorate Evaluation Studies*. Dublin: The Government Stationery Office.

Department of Education and Science and ICT Ireland (2009). *Smart Schools = Smart Economy*. Report of the ICT in Schools Joint Advisory Group to the Minister for Education and Science. Ireland: DES.

Department of Education and Skills (2010). Better Literacy and Numeracy for Children and Young People. A Draft National Plan to Improve Literacy and Numeracy in Schools. Dublin: The Government Stationery Office.

Department of Education and Skills (2011). *Literacy and Numeracy for Learning and Life. The National Strategy to Improve Literacy and Numeracy among Children and Young People 2011-2020.* Dublin: Department of Education and Skills. Available online at: http://www.education.ie/admin/servlet/blobservlet/lit_num_strat.pdf?language=EN&igstat=true. [Accessed: 14/07/12].

Department of Education and Skills (2012). *Minister Quinn announces details of 275 major school building projects - More than 15,000 jobs to be created over five years*. Government Press Release, 12 March 2012. Available online at: http://www.education.ie/en/Press-Events/Press-Releases/2012-Press-Releases/PR12-03-12.html [Accessed: 25/11/12].

Department of Education and Skills (2012a). Textbook Rental Schemes in Schools and the Allocation of Textbook Grants by the Department of Education and Skills. A Report for Consultation. Ireland: DES. Available online: http://www.education.ie/en/Schools-Colleges/Services/Grants-and-Additional-Support/schools_book_rental_scheme_report.pdf [Accessed: 02/12/12].

Department of Education and Skills (2012b). *Guidelines for Developing Textbook Rental Schemes in Schools. Draft for Consultation.* Ireland: DES. Available online: http://www.education.ie/en/Schools-Colleges/Services/Grants-and-Additional-Support/schools_book_rental_scheme_dft_guide.pdf [Accessed: 02/12/12].

Department of Education and Skills (2012c, May 28). 28 May 2012: New Department of Education & Skills survey shows that more than three quarters of primary schools operate book rental schemes - Minister Quinn announces he has ring-fenced €15m in book grants for the next school year. DES Press Release. Available online: http://www.education.ie/en/Press-Events/Press-Releases/2012-Press-Releases/PR12-05-28.html [Accessed: 02/12/12].

Department of Finance, (2012). Summary of 2012 Budget and Estimate Measures. Policy Changes. Available online at: http://budget.gov.ie/Budgets/2012/Documents/Summary%20of%202012%20Budget%20and%20Estimates%20Measures%20Policy%20Changes.pdf [Accessed: 26/09/2012]

Design-Based Research Collective (DBRC) (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher 32*(1), 5-8.

Díaz-Negrillo, A. & Fernández-Dominguez, J. (2006) Error Tagging Systems for Learner Corpora. *RESLA 19*, pp. 83-102.

Donovan, P. (1998). Piloting: A Publisher's View. In B. Tomlinson (Ed.) *Materials Development in Language Teaching*. Cambridge: Cambridge University Press, pp. 149-189.

Durrell, M., Bennett, P., Scheible S. and Whitt, R. J. (2012). *The GerManC Corpus*. Manchester: University of Manchester. Available online at: http://www.alc.manchester.ac.uk/subjects/german/research/projects/germanc/germancplus/ [Accessed: 21/03/15].

Education Services Australia, (2012). *Educational value for standard digital resources. Version 1.0.* Australia: Education Services Australia. Available online at: http://www.ndlrn.edu.au/verve/_resources/Educational_value_standard_for_digital_resources.pdf#search=educational%20value%20standard%20for%20digital%20resources [Accessed: 23/04/13].]

Eivers, E., Shiel, G., Perkins, R. and Cosgrove, J. (2005). *Succeeding in Reading? Reading standards in Irish primary schools.* Report prepared for the DES by the Educational Research Centre (ERC). Dublin: Educational Research Centre.

Ellis, N. (2006). Cognitive perspectives on SLA. The Associative-Cognitive CREED. *AILA Review 19 (2006)*, pp. 100-121.

Ellis, R. (1997). *Second Language Acquisition.* Oxford, New York: Oxford University Press.

European Commission, (2013). Survey of Schools: ICT in Education. Benchmarking access, use and attitudes to technology in Europe's schools. Final Study Report. Belgium: European Union. Available online at: https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/KK-31-13-401-EN-N.pdf [Accessed: 15/07/14].

Farr, F., A. Chambers & S. O'Riordan (2010). Corpora for materials development in language teacher education: Underlying principles and useful data. In F. Mishan & A. Chambers (Eds.), *Perspectives on Language Learning Materials Development*. Bern: Peter Lang, pp. 33–62.

Fogarty, R. (1991) *The Mindful School: How to Integrate the Curricula*. Palatine, IL: Skylight Publishing, Inc.

Fox, G. (1998). Using corpus data in the classroom. In B. Tomlinson (Ed.) *Materials Development in Language Teaching.* Cambridge: Cambridge University Press, pp. 25-43.

Frazier, J. R. and Juza, P. (2008). Materials Used in the USA. In B. Tomlinson (Ed.). *English Language Learning Materials: A Critical Review*. London, New York: Continuum. pp. 179-192.

Gamper, J. and Knapp, J. (2002). XML for an electronic learners' dictionary. Proceedings of the IADIS International Conference WWW/Internet 2002, ICWI 2002, Lisbon, Portugal, November 13-15, 2002. IADIS 2002. pp. 427-435. Available online at: http://www.informatik.uni-trier.de/~ley/db/conf/iadis/icwi2002.html [Accessed: 18/07/14].

Gilleece, L., Shiel, G., Clerkin, A. and Millar, D. (2012). *The 2010 National Assessments of English Reading and Mathematics in Irish-Medium Schools. Main Report*. Dublin: Educational Research Centre. Available online at: http://www.erc.ie/documents/naims_mainreport.pdf [Accessed: 25/11/12].

Girard, D., Courtillon, J., Page, B. and Richterich, R. (1988). *Selection and distribution of contents in language syllabuses*. Strasbourg: Council of Europe.

Gottheim, L. (2010). Composing Textbooks as a Non-Expert. In B. Tomlinson and H. Masuhara (Eds.) *Research for Materials Development in Language Learning: Evidence for Best Practice.* London, New York: Continuum, pp. 224-236.

Government of Ireland, (1999). *Primary School Curriculum.* Dublin: The Government Stationery Office.

Granger, S. (2003). Error-tagged learner corpora and CALL: a promising synergy. *CALICO Journal 20*(3), pp.465-480.

Granger, S., Kraif, O., Ponton, C., Antoniadis, G. and Zampa, V. (2007). Integrating Learner Corpora and Natural Langauge Processing: A Crucial Step Ttowards recomciling technological sophistication and pedagogical effectiveness. *ReCALL 19(3).* pp. 252-268. UK: Cambridge University Press.

Gray, J. (2013). Introduction. In J. Gray (Ed.) *Critical Perspectives on Language Teaching Materials*. UK, USA: Palgrave, Macmillan.

Greene, C. (2013). *Computer assisted (language) learning (CA(L)L) for the inclusive classroom*. Unpublished PhD Thesis. Dublin City University, Dublin.

Hann, N., Timmis, I. and Masuhara, H. (2010). ESOL Materials: Practice and principles. In F. Mishan & A. Chambers (Eds.), *Perspectives on Language Learning Materials Development*. Bern: Peter Lang, pp. 223-248.

Harbison, L. B. (2008). The Utility of Textbooks in the Teaching and Learning of Mathematics with Understanding in Classrooms of the Lower Primary Years in Ireland. Unpublished Ph.D. thesis, University of Newcastle, UK.

Harris, J. and O'Leary, D. (2009). A Third Language at Primary Level in Ireland. In M. Nikolov (Ed.) *Early Learning of Modern Foreign Languages. Processes and Outcomes.* Bristol, Buffalo, Toronto: Multilingual Matters, pp. 1-13.

Harris, J. and O'Leary, D. (2012). A Third Language at Primary Level in Ireland: An independent evaluation of the Modern Languages in Primary Schools Initiative*:* In *Final Report on the Modern Languages in Primary Schools Initiative 1998-2012.* Kildare: Kildare Education Centre, pp. 39-50. Available online at: http://www.mlpsi.ie [Accessed: 28/09/2012]

Harris, J., Forde, P., Archer, P., Nic Fhearaile, S. and O'Gorman, M. (2006). *Irish in Primary Schools: Long-Term National Trends in Achievement*. Dublin: Department of Education and Science.

Harris, J. and Conway, M. (2002). Modern Languages in Primary Schools: An Evaluation of the National Pilot Project. Dublin: Institiúid Teangeolaíochta Éireann.

Harwood, N. (2010). Issues in materials development and design. In N. Harwood (Ed.) *English Language Teaching Materials: Theory and Practice.* New York: Cambridge University Press, pp. 3-32.

Hawkins, E. W. (1974). Modern Languages in the Curriculum. In G. Perren (Ed.) *The Space Between: English and Foreign Languages at School*. London: CILT.

Hawkins, E. W. (1999). Foreign Language Study and Language Awareness. *Language Awareness, 8*(3-4), pp. 124-142.

Heift, T. (1998). *Designed Intelligence: A Language Teacher Model.* Unpublished PhD Thesis, Simon Fraser University, Burnaby.

Heift, T. and Schulze, M. (2007). *Errors and Intelligence in Computer-Assisted Language Learning: Parsers and Pedagogues.* New York, London: Routledge.

Hémard, D. (2004). Enhancing Online CALL Design: The Case for Evaluation. *ReCALL, 16*(2), pp. 502-519.

Heughebaert, E. and Troubleyn, K. (1996). Company-Specific Language Courseware: Bellingua and Linguatex as case studies. *Computer Assisted Language Learning, 9*(4). pp. 349-358.

Hubbard, P. L. (1996). Elements of CALL Methodology: Development, Evaluation, and Implementation. In M. C. Pennington (Ed.) *The Power of CALL*. Houston, TX: Athelstan.

Hubbard, P. (206). Evaluating CALL software. In: L. Ducate and N. Arnold (Eds.), *Calling on CALL: From Theory and Research to New Directions in Foreign Language Teaching.* San Marcos, TX: CALICO, pp. 313-334.

Hutchins, W. J. and Somers, H. L. (1992). *An Introduction to Machine Translation*. London: Academic Press.

Hutchinson, T. and Waters, A. (1987). *English for Specific Purposes: A Learning-Centred Approach.* Cambridge: Cambridge University Press.

INCA, (2013). *Control and Supply of School Textbooks: INCA Comparative Tables.* Available online at: http://www.nfer.ac.uk/what-we-do/information-and-reviews/inca/INCAcomparativetablesMarch2012.pdf [Accessed: 17/06/14].

Irish Primary Principals' Network (2008). Tesco exploits seven years of neglect by Government of ICT funding for schools. In *Irish Primary Principals' Network* website. Available online at: www.ippn.ie [Accessed: 10/06/08].

Irish National Teachers' Organisation (INTO) (1991). Foreign Language Teaching in Primary Schools: Issues and Research. Dublin: Author.

Irish National Teachers' Organisation (2008). The Primary School Curriculum. INTO Survey 2005 and Proceedings of the Consultative Conference on Education 2006.

Jalkanen, J. (2013). Dynamics of sustainable pedagogical development: Insights from higher education language teaching. In *Conference proceedings from WorldCall 2013: Global perspectives on Computer-Assisted Language Learning.* Glasgow, UK: University of Ulster, pp. 121-123.

James, C. (1998). *Errors in Language Learning and Use: Exploring Error Analysis*. Edinburgh: Pearson Education Limited.

Jolly, D. and Bolitho, R. (2011) A Framework for Materials Writing. In B. Tomlinson (Ed.) *Materials Development in Language Teaching (2nd Ed.)*. Cambridge: Cambridge University Press, pp. 107-134.

Kaiser, H. (2013). *A Close Look at Simplified Technical English*. E-Magazine: TCWorld: Technical Communication. Magazine for International Information Management. September 2013 Available online http://www.tcworld.info/e-magazine/technical-communication/article/a-close-look-at-simplified-technical-english [Accessed: 11/11/'13].]

Keogh, K. (2011). Using Mobile Phones for Teaching, Learning and Assessing Irish in Ireland: Processes, Benefits and Challenges. In Ng (Ed.) Mobile Technologies and Handheld Devices for Ubiquitous Learning: Research and Pedagogy. Hershey, New York: Information Science Reference. pp. 209-236.

Keogh, K., Koller, T., Ward, M., Uí Dhonnchadha, E. & van Genabith, J. (2004) CL for CALL in the Primary School. In *Proc. of the Workshop eLearning for Computational Linguistics and Computational Linguistics for eLearning*, pp. 79-85, Geneva, Switzerland.

Kildare Education Centre, (2012). *Final Report on the Modern Languages in Primary Schools Initiative 1998-2012.* Available online at: http://www.mlpsi.ie [Accessed: 28/09/2012]

Kittredge, R. I. (2003). Sublanguages and Controlled Languages. *In*: R. Mitkov (Ed.) *The Oxford Handbook of Computational Linguistics.* New York: Oxford University Press.

Knapp, J. (2004). *A New Approach to CALL Content Authoring*. Unpublished Ph.D. thesis, Fachbereich Informatik der Universitaet Hannover, Germany.

Koizumi, R. (2012). Relationships Between Text Length and Lexical Diversity Measures: Can We Use Short Texts of Less than 100 Tokens? *Vocabulary Learning and Instruction, 1(1),* pp. 60-69. Available online at: http://vli-journal.org/wp/vli-v01-1-2187-2759/ [Accessed: 11/11/14].

L'Haire, S. (2007). FipsOrtho: A spell checker for learners of French. In *ReCALL, 19(2),* pp 137-161.

Lake, K. (1994). Integrated Curriculum. Close-Up #16. In *School Improvement Research Series, NWREL*. Available online at: http://educationnorthwest.org/webfm_send/528 [Accessed: 27/10/11].

Leacock, T. L. and Nesbit, J. C. (2007). A Framework for Evaluating the Quality of Multimedia Learning Resources. *Educational Technology & Society, 10 (2),* pp. 44-59.

Leech, G. (2005). Adding Linguistic Annotation. In M. Wynne (Ed.) *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books. Pp. 17-29. Available online from http://ahds.ac.uk/linguistic-corpora/ [Accessed: 29/11/14].

Lightbown, P. M. and Spada, N. (2006). *How Languages are Learned* (3rd Ed.). Oxford: Oxford University Press.

Liou, H., Chang, M., Chao, Y. J., Wu, W. V., Lee, H. and Tsai, Y. (2013). Social networking, collaborative learning, and emerging technologies: Sustainability claimed and projected. In *Conference proceedings from WorldCall 2013: Global perspectives on Computer-Assisted Language Learning.* Glasgow, UK: University of Ulster, pp. 194-199.

Littlejohn, A. (2011). The analysis of language teaching materials: Inside the Trojan Horse. In B. Tomlinson (Ed.) *Materials Development in Language Teaching (2nd Ed.)*. Cambridge: Cambridge University Press, pp. 179-211.

Littlejohn, A. (2012). Language Teaching Materials and the (very) big Picture. *Electronic Journal of Foreign Language Teaching, 9(1)*: p. 283-297.

Levy, M. and Kennedy, C. (2010). Materials Development in Three Italian CALL Projects: Seeking an Optimal Mix Between In-Class and Out-Of-Class Learning. *CALICO Journal 27*(3), pp. 529-539.

Lumala, M. and Trabelski, S. (2008). Materials Used in Africa. In B. Tomlinson (Ed.) *English Language Learning Materials: A Critical Review.*. London, New York: Continuum. pp. 223-244.

Mackey, A., Gass, S. and McDonagh, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition 22(4)*, pp. 471-497.

Maley, A. (2011). Squaring the circle – reconciling materials as constraint with materials as empowerment. In B. Tomlinson (Ed.) *Materials Development in Language Teaching (2nd Ed.)*. Cambridge: Cambridge University Press, pp. 379-402.

Maley, A. (2013). Creative Approaches to Writing Materials. In Tomlinson, B. (Ed.) *Developing Materials for Language Teaching (2nd Ed.)*. London, New York: Continuum. pp. 167-188.

Malvern, D., Richards, B., Chipere, N. and Durán, P. (2004). Lexical diversity and language development: Quantification and assessment. Hampshire, NY: Palgrave Macmillan.

Mares, C. (2003). Writing a Coursebook. In B. Tomlinson (Ed.) *Developing Materials for Language Teaching.* London, New York: Continuum. pp. 130-140.

McCarten, J. and McCarthy, M. (2010). Bridging the gap between corpus and course book: The case of conversation strategies. In F. Mishan & A. Chambers (Eds.), *Perspectives on language learning materials development*. Bern: Peter Lang, p.11-32.

McCoy, S., Smyth, E., and Banks, J. (2012). *Learning in Focus: The Primary Classroom: Insights from the Growing Up in Ireland Study.* Additional investigation of ESRI data commissioned by the National Council for Curriculum and Assessment. Available online at: http://www.ncca.ie/en/The_Primary_Classroom_ESRI_January_18_2012.pdf [Accessed: 22/09/12].

McEnery, A. M. (1992). *Computational Linguistics – a Handbook and Toolbox or Natural Language Processing*. UK: Sigma Press.

McEnery, T. and Xiao, R. (2010). What Corpora Can Offer in Language Teaching and Learning. In E. Hinkel (Ed.) *Handbook of Research in Second Language Teaching and Learning Volume II.* New York, London: Routledge Taylor and Francis Group. *p. 364-380.*

McGrath, I. (2002). *Materials Evaluation and Design for Language Teaching.* Edinburgh: Edinburgh University Press.

McKay, P. (2006). *Assessing Young Language Learners.* Cambridge: Cambridge University Press.

Microsoft (2014). *Test your document's readability.* Microsoft online support. Available online at: http://office.microsoft.com/en-ie/word-help/test-your-document-s-readability-HP010148506.aspx [Accessed: 11/11/14].

Millett, A. & Johnson, D. (1996). Solving teachers' problems? The role of the commercial mathematics scheme. In Johnson & Millett *Implementing the Mathematics National Curriculum. Policy, Politics and Practice*, pp.54-74. Paul Chapman Publishing: London.

Mukundan, J. (2010). Words as they appear in Malaysian Secondary School English Language Textbooks: Some implications for pedagogy. In B. Tomlinson and H. Masuhara (Eds.). *Research for Materials Development in Language Learning. Evidence for Best Practice.* London, NY: Continuum. pp. 291-304.

National Council for Curriculum and Assessment, (1999). Pilot Project on Modern Languages in the Primary School: Draft Curriculum Guidelines. Dublin: Author.

National Council for Curriculum and Assessment, (2001). *Modern Languages in Primary Schools. Teacher Guidelines*. Dublin: Author.

National Council for Curriculum and Assessment, (2001a). *Specifications for Publishers: English*. Dublin: Author.

National Council for Curriculum and Assessment, (2004). Curriculum, Assessment and ICT in the Irish Context: A Discussion Paper. Dublin: Author.

National Council for Curriculum and Assessment, (2004a). Information and Communications Technology (ICT) in the Primary School Curriculum. Guidelines for Teachers. Dublin: Author.

National Council for Curriculum and Assessment, (2005). Report on the Feasibility of Modern Languages in the Primary School Curriculum. Dublin: Author.

National Council for Curriculum and Assessment, (2005a). Primary Curriculum Review Phase One: Summary of Findings and Recommendations. Dublin: Author.

National Council for Curriculum and Assessment, (2005b). *Primary School Curriculum: English. Additional Support Materials*. Dublin: Author.

National Council for Curriculum and Assessment, (2007). ICT Framework: A Structured Approach to ICT in Curriculum and Assessment. A report on the School-based Developmental Initiative. Dublin: Author.

National Council for Curriculum and Assessment, (2008). Modern Languages in the Primary School Curriculum: Feasibility and Futures Report. Dublin: Author.

National Council for Curriculum and Assessment, (2008a). *Primary Curriculum Review Phase Two.* Dublin: Author.

National Council for Curriculum and Assessment (2009). *Aistear, the Early Childhood Curriculum Framework*. Available online at: http://www.ncca.biz/Aistear/ [Accessed: 22/01/10].

National Council for Curriculum and Assessment, (2010). Curriculum Overload in Primary Schools: Experiences and reflections from the Learning Site. Dublin: Author.

National Council for Curriculum and Assessment, (2010a). Curriculum Overload in Primary Schools: An overview of national and international experiences. Dublin: Author.

National Council for Curriculum and Assessment, (2010b). *Better Literacy and Numeracy for Children and Young People*. NCCA Submission. Dublin: Author

National Council for Curriculum and Assessment, (2012). *Priorities for Primary Education? Report on Responses to "Have your Say"*. Dublin: Author.

National Council for Curriculum and Assessment, (2014). *Consultation on the Draft Primary Language Curriculum, Junior Infants to Second Class. Interim report*. Dublin: Author.

National Policy Advisory and Development Committee (2001). *The Impact of schools IT2000: Report and Recommendations to the Minister for Education and Science.* Dublin: National

Policy Advisory and Development Committee. Available online at: http://ww.ncte.ie/npadc/ [Accessed: 31/07/2002].

NFER, (1996). *Thematic Probe: Control and Supply of Textbooks.* International Review of Curriculum and Assessment Frameworks. UK: Le Métais, J.

Nunan, D. (1988). Principles for designing language teaching materials. *Guidelines 10*(2), pp. 1-24.

O'Brien, S. (2010). Controlled language and readability. *Ata Scholarly Monograph Series*, 15, pp. 143-168.

Ó'Duibhir, P. and Cummins, J. (2012). *Towards an Integrated Language Curriculum in Early Childhood and Primary Education* (3-12 years). NCCA Research Report No. 16. Dublin: NCCA.

OECD, (2012). *Education at a Glance 2012.* Paris: OECD Publishing.

OECD, (2013). *Innovative Learning Environments*. Educational Research and Innovation, OECD Publishing. Available online at: http://www.keepeek.com/Digital-Asset-Management/oecd/education/innovative-learning-environments_9789264203488-en#page4 [Accessed: 18/07/14].

OECD, (2013a). *PISA 2012 Results. What makes schools successful?Resources, policies and practices. (Volume IV).* PISA: OECD Publishing.

OECD, (2014). *Measuring Innovation in Education.* Educational Research and Innovation, OECD Publishing. Available online at: http://www.keepeek.com/Digital-Asset-Management/oecd/education/measuring-innovation-in-education_9789264215696-en#page1 [Accessed: 18/07/14].

Oster, L. (1993). Sub-Saharan Africa: An Interdisciplinary Curriculum Unit. *English Journal 82/4:* pp. 24-28.

Palmer, H. E. (1917). *The scientific study and teaching of languages.* London: Harrap.

Palmer, J. (1991). Planning Wheels Turn Curriculum Around. *Educational Leadership 49/2*: pp. 57-60.

Perkins, R., Moran, G., Cosgrove, J. and Shiel, G. (2010). *PISA 2009: The Performance and Progress of 15-Year Olds in Ireland. Summary Report*. Dublin: ERC.

Perkins, R., Shiel, G., Merriman, Cosgrove, J. and Moran, G. (2013). *Learning for Life: The Achievements of 15-year-olds in Ireland on Mathematics, Reading Literacy and Science in PISA 2012.* Dublin: ERC.

Pienemann, M. (2008). A Brief Introduction to Processability Theory. In J.-U. Keβler (Ed.) *Processability Approaches to Second Language Development and Second Language Learning*. Newcastle: Cambridge Scholars Publishing, pp. 9-30.

Popovici, R. and Bolitho, R. (2003). The Romanian Textbook Project. In B. Tomlinson (Ed.) *Developing Materials for Language Teachers.* London, New York: Continuum. pp. 505-517.

Prabhu, N.S. (1988) Materials as Support; Materials as Constraint. RELC Working Papers on Materials for Language Learning and Teaching: Trends and Developments, 1 (April) (Singapore: SEAMEO).

Prodromou, L. and Mishan, F. (2008). Materials Used in Western Europe. In B. Tomlinson (Ed.) *English Language Learning Materials: A Critical Review*. London, New York: Continuum. pp. 193-212.

Prowse, P. (2011). How writers write: Testimony from authors. In B. Tomlinson (Ed.) *Materials Development in Language Teaching (2nd Ed.)*. Cambridge: Cambridge University Press, pp. 151-173.

Rimrott, A. and Heift, T. (2005). Language learners and generic spell-checkers in CALL. *CALICO Journal, 23(1),* pp. 17-48.

Royal Irish Academy (2011). National Languages Strategy: Royal Irish Academy National Committee for Modern Language, Literacy and Cultural Studies. Dublin: Royal Irish Academy. Available online at: www.ria.ie [Accessed: 02/12/11].

Ryan, L., Anastasiou, D. and Cleary, Y. (2009). Using content development guidelines to reduce the cost of localising digital content. *Localisation Focus* 8(1), pp. 11-28.

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging using Decision Trees. *Proceedings of the ACL SIGDAT-Workshop* 1994. Dublin, Ireland.

Scott, M., Carioni, L., Zanatta, M., Bayer, E. and Quintanilha, T. (1984). Using a "standard exercise" in teaching reading comprehension. *ELT Journal 38*(2), pp. 114-120

Selinker, L. (1972). Interlanguage. *IRAL - International Review of Applied Linguistics in Language Teaching.10* (3), pp. 209-232.

Selinker, L. (1992). *Rediscovering Interlanguage.* London: Longman.

Shavelson, R. J., Phillips, D. C., Towne, L. & Feuer, M. J. (2003). On the Science of Educational Design Studies. *Educational Researcher*, *32*(1), 25-28.

Schulze, M. (2003). Grammatical Errors and Feedback: Some Theoretical Insights. *CALICO Journal 20*(3), pp. 437-450.

Shiel, G. and O'Flaherty, A. (2006). *NCTE 2005 Census on ICT Infrastructure in Schools. Statistical Report.* Dublin: NCTE; Available online at: http://test.ncte.ie/census/Census2005/ [Accessed: 04/07/06]

Shoemaker, B. (1989). Integrative Education: A Curriculum for the Twenty-First Century. In *Oregon School Study Council 33/2*.

Shen, H., Yuan, Y. and Ewing, R. (2015) English learning websites and digital resources from the perspective of Chinese university EFL practitioners. Cambridge University Press: *ReCALL* Firstview article, pp. 1-21.

Sinclair, J. M. (1986). Basic computer processing of long texts. In G. Leech & C. N. Candin (Eds.), *Computers in English Language Teaching and Research.* London: Longman.

Smyth, E., Dunne, A., McCoy, S. and Darmody, M. (2006). Pathways through the Junior Cycle: The Experiences of Second Year Students. ESRI: NCCA and the Gender Equality Unit funded research report. Dublin: The Liffey Press in association with the Economic and Social Research Institute.

Smyth, E., Darmody, M., McGinnity, F., & Byrne, D. (2009). *Adapting to diversity: Irish schools and newcomer students.* Dublin: The Economic and Social Research Institute.

Sommerville, I. (2004). *Software Engineering*. Harlow, Essex, England, New York: Pearson/Addison Wesley. 7th Edition.

Spada, N. & Fröhlich, M. (1995). *The Communicative Orientation of Language Teaching Observation Scheme: Coding Conventions and Applications.* Sydney: National Centre for English Language Teaching and Research, Macquarie University.

Svalberg, A. M. L. (2001). The Problem of False Language Awareness. *Language Awareness* 10 (2&3), pp. 200-212.

Swan, M. and Walter, C., (1984). *The Cambridge English course, book 1*. Cambridge: Cambridge University Press.

Thornbury, S. (2013). Resisting Coursebooks. In J. Gray (Ed.) *Critical Perspectives on Language Teaching Materials.* Hampshire, New York: Palgrave Macmillan. pp. 204-223.

Thouësny, S. (2011). *Modelling second language learners' interlanguage and its variability: A computer-based dynamic assessment approach to distinguishing between errors and mistake.* Unpublished Ph.D. thesis, Dublin City University, Ireland.

Tomlinson, B. (1999). Developing criteria for evaluating L2 materials. *IATEFL Issues, Feb.-Mar. 1999*, pp. 10-13.

Tomlinson, B. (2000). Beginning to Read Forever. *Reading in a Foreign Language 13*(1), pp. 523-538.

Tomlinson, B. (2008). Language Acquisition and Language Learning Materials. In B. Tomlinson (Ed.) *English Language Learning Materials*. London, New York: Continuum. pp. 3-13.

Tomlinson, B. (2011). Introduction: Principles and procedures of materials development. In B. Tomlinson (Ed.) *Materials Development in Language Teaching (2nd Ed.)*. Cambridge: Cambridge University Press, pp. 1-34.

Tomlinson, B. (2012). Materials Development for Language Learning and Teaching. *Language Teaching*, 45(2), pp 143-179.

Tomlinson, B. (2013). Developing Principled Frameworks for Materials Development. In B. Tomlinson (Ed.) *Developing Materials for Language Teachers (2nd Ed.).* London, New York: Continuum. pp. 95-118.

Tomlinson, B. (2013a). Comments on Part E. In B. Tomlinson (Ed.) *Developing Materials for Language Teachers (2nd Ed.).* London, New York: Continuum. pp. 537-540.

Uí Dhonnchadha, E. (2002). *An Analyser and Generator for Irish Inflectional Morphology Using Finite-State Transducers*. MSc Thesis, Dublin City University, Ireland

Van Elsen, E., van Deun, K. and Decoo, W. (1991). Wordchip: The application of external versatility to an English lexical CALL program. *System 19*(4). pp. 401-417.

Van Genabith, J. (2009). Next generation localisation. *Localisation Focus* 8(1), pp. 4-10

Vandeventer, A. (2001). Creating a Grammar Checker for CALL by Constraint Relaxation: A Feasibility Study. *ReCALL, 13(1),* pp. 110-120.

Varley, Murphy and Veale (2008). Science in Primary Schools, Phase 1, Final Report.

Vercellotti, M. L. (2012). *Complexity, Accuracy, and Fluency as Properties of Language Performance: The Development of the Multiple Subsystems over Time and in Relation to Each Other*. Unpublished PhD Thesis, University of Pittsburgh, Pittsburgh. Available online at: http://d-scholarship.pitt.edu/12071/ [Accessed: 11/11/'14].

Vygotsky, L. S., (1978). *Mind in Society*. Cambridge MA: Harvard University Press.

Walz, J. (1982). Error correction techniques for the foreign language classroom. In *Language in education: Theory and practice series,* 50. Washington D. C.

Ward, M. (2002). Reusable XML Technologies and the Development of Language Learning Materials. *ReCALL, 14(2),* pp. 283-292.

Ward, M. (2002a). *A Template for CALL Programs for Endangered Languages,* Unpublished M.Sc. thesis. Dublin City University. Available online at: http://www.computing.dcu.ie/~mward/mthesis.html [Accessed: 06/01/13].

Ward, M. (2013). Factors for sustainable CALL. In *Conference proceedings from WorldCall 2013: Global perspectives on Computer-Assisted Language Learning.* Glasgow, UK: University of Ulster, pp. 357-359.

Waring, R. and Nation, P. (2004). Second language reading and incidental vocabulary learning. In D. Albrechtsen, K. Haastrup & B. Henriksen (Eds.), *Writing and vocabulary in foreign language acquisition. Serie: Angles on the English-speaking world* (Vol. 4, pp. 97-110). Copenhagen: Museum Tusculanum Press.

Wasburn, L. H. (1997). Accounts of Slavery: An Analysis of United States history textbooks from 1900 to 1992. *Theory and Research in Social Education 21*(3), pp. 232-256.

West, M. (1953). *A general service list of English words*. London: Longman.

Willis, J. (2011). Concordances in the classroom without a computer: Assembling and exploiting concordances of common words. In B. Tomlinson (Ed.) *Materials Development in Language Teaching (2nd Ed.).* Cambridge: Cambridge University Press, pp. 51-77.

Wilson, B. G. (1996). What is a Constructivist Learning Environment? In B. G. Wilson (Ed.) *Constructivist Learning Environments: Case Studies in Instructional Design.* New Jersey: Educational technology Publications.

Wolfe-Quintero, K., Inagaki, S. and Kim, H. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy and Complexity*. USA: University of Hawaii Press.

Yule, G. (2010). *The Study of Language* (4th Edition). Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo, Delhi, Mexico City: Cambridge University Press.

Zilversmit, A. (1993). *Changing Schools: Progressive Education Theory and Practice, 1030-1960*. Chicago: University of Chicago Press

APPENDICES

Appendix A Summary of principle-based approaches to materials development

Source	Context	Principles outlined
Farr, Chambers and O'Riordan (2010)	Materials for language teacher education	1. Materials should achieve impact 2. Materials should help learners to feel at ease and to develop confidence 3. Materials should require and facilitate critical cognitive and emotional development 4. Materials should allow for attention to be drawn to specific features 5. Materials should permit a silent period and take into account that the positive effects of instruction are usually delayed
Hann, Timmis and Masuhara (2010)	Learner-related principles for English for Speakers of Other Languages (ESOL)	1. ESOL materials should cater for learners' real and immediate needs 2. ESOL materials should provide cognitive and effective engagement in addition to catering for learners' real and immediate needs 3. ESOL materials should respect and exploit the diversity of ESOL learners' backgrounds 4. ESOL materials for beginners should support the learning of learners with little or no literacy skills 5. ESOL materials should be developed by learners aswell 6. ESOL materials should optimise learner development i.e. improve skills for learning 7. ESOL materials should be equally accessible to learners and teachers 8. ESOL materials should be suited to both high tech and low tech environments 9. While linked to the ESOL core curriculum, materials should be flexible enough for roll-on/roll-off enrolment.
Tomlinson (2011)	Basic principles of SLA relevant to materials development	1. Materials should achieve impact 2. Materials should help learners to feel at ease 3. Materials should help learners to develop confidence 4. What is being taught should be perceived by learners as relevant and useful 5. Materials should require and facilitate learner self-investment 6. Learners must be ready to acquire the points being taught 7. Materials should expose the learners to language in authentic use 8. The learners' attention should be drawn to linguistic features of the input 9. Materials should provide the learners with opportunities to use the target language to achieve communicative purposes 10. Materials should take into account that the positive effects of instruction are usually delayed

Source	Context	Principles outlined
		11. Materials should take into account that learners differ in learning styles 12. Materials should take into account that learners differ in affective attitudes 13. Materials should permit a silent period at the beginning of instruction 14. Materials should maximise learning potential by encouraging intellectual, aesthetic and emotional involvement which stimulates both right and left brain activities 15. Materials should not rely too much on controlled practice 16. Materials should provide opportunities for outcome feedback.
Bell and Gower (2011)	Guiding principles for their own materials development work	1. Flexibility: the materials should facilitate the teacher in using them flexibly 2. From text to language: Authentic texts at an appropriate level 3. Engaging content: content to try to engage the learner personally 4. Natural language: Spoken texts were to be authentic with 'real' speakers rather than actors. 5. Analytic Approaches: Students to deduct their own rules for grammar 6. Emphasis on review: Presented grammar in review format with the assumption that their intermediate students already had an existing knowledge of the topics 7. Personalised practice: Learners practice sentences about themselves (e.g. I would like to become…) 8. Integrated skills: Four skills should be presented in an integrated way 9. Balance of approaches 10. Learner development 11. Professional respect for the finished product.
Hutchinson and Waters (1987) as cited in McGrath (2002)	Guiding principles as experienced materials writers	1. Use existing materials as sources of ideas 2. Work with other people if possible 3. Don't expect to write materials that are perfect first time 4. Don't underestimate the time needed 5. Pay attention to the appearance of the materials.
Nunan (1988) as cited in McGrath (2002)	Guiding principles as an experienced materials writer	1. Materials should be clearly linked to the curriculum they serve 2. Materials should be authentic in terms of text and task 3. Materials should stimulate interaction 4. Materials should allow learners to focus on formal aspects of the language 5. Materials should encourage learners to develop learning skills, and skills in learning 6. Materials should encourage learners to apply their developing language skills to the world beyond the classroom.

Table A.1 Summary of principle-based approaches to materials development

Appendix B Informed Consent Form for Parents/Guardians and Students

Dear Parent/Guardian,

A. As part of your child's German course, they will be producing written work and completing written and oral tests. I would like your permission to record their oral tests and analyse their written work and oral test to help me decide what materials may need to be revised and to feed into my research at DCU.

B. Your child will also be using computer software to help them learn German. I would like to monitor what they do while using the software and have them fill in evaluation questionnaires afterwards.

C. We'll also have discussions about what they do or do not like learning and whether they enjoy learning with the computer.

All data gathered will be use to improve the quality of your child's German course and the software being developed to help them to learn German as part of a research project. More detailed information on this research project can be found on the attached Plain Language Statement. If you decide not to consent to A, B and C above, your child can still learn German with the class but none of your child's data (see A, B, and C above) will be gathered for the project.

Thank you,
Miss. Keogh

---- 8< ----------*Please detach and return*----------

A. ☐ I agree that my child _____ may be recorded during his/her oral test and that the work he/she produces may be analysed.

B. ☐ I agree that my child (named in A above) may be monitored while he/she is using the German software and that he/she may complete a questionnaire about what he/she thinks about using the software.

C. ☐ I agree that he/she can participate in class discussions and that information gathered may be used for the research project.

(Please note that I must inform you that the confidentiality of information provided can only be protected within the limitations of the law).

Signed: _____ **Date**: _____

Parent/Guardian

Signed: _____ **Date**: _____

Student

Appendix C Listing of the relevant texts and electronic language learning resources consulted in the development of the raw linguistic content

Textbooks

Büttner, S., Kopp, G. & Alberti, J. (1996). *Tamburin. Deutsch für Kinder*. Ismaning: Max Hueber Verlag.

Moeller, J. (2010). Deutsch heute: introductory German. Boston, MA: Cengage Learning.

Sutton, S. (1993). "*Sagen Sie mal!" A course in oral German for the leaving certificate*. Tallaght, Dublin: Folens Publishing.

Websites

http://gut.languageskills.co.uk/

http://www.bbc.co.uk/languages/german/index.shtml

http://class.georgiasouthern.edu/german/grammar/grammar.htm

http://dict.leo.org/

http://www.washjeff.edu/capl/

http://www.2rhyme.ch/Search/Default.asp?WordEnding=ichten&Submit=Reimen

http://www.iee.et.tu-dresden.de/cgi-bin/cgiwrap/wernerr/search.sh

http://www.vokabel.com/

http://www.languageguide.org/deutsch/index.jsp

http://www.juma.de/v.php?fl=welcome.html

http://www.colby.edu/german/deutsch_ueben/

http://www.aatg.org/

http://www.uq.net.au/%7Ezzkmunr1/german/index.htm

http://www.uncg.edu/%7Elixlpurc/publications/NetzUeb.html

CD Roms

- Vocabulary Builder German
- EuroTalk interactive series (Vocabulary Builder, Talk Now Beginners, World Talk Intermediate and Movie Talk)
- Learning German with a Cartoon disc for 4-12 year olds

Appendix D Raw linguistic content version 1.0

Content following initial drafting of the raw linguistic resources and consultation of alternative resources in Phase I.

Note: Highlighted items were added from the review of alternative resources.

Section One: General Introductions

Language:

1. Greetings Hallo! Guten Morgen Guten Tag Guten Abend Gute Nacht Schlaf gut	 Auf Wiedersehen Tschüs Bis bald Bis morgen/Donnerstag…
2. How are you? Wie geht's? Gut, danke! Sehr gut, danke! Nicht so gut Schlecht Mir geht's gut/sehr gut…	3. What is your name? Wie heißt du? Ich heiße _____

Section Two: Parts of the body

Language:

Der Bauch Der Daumen Der Finger Der Fuß Der Hals Der Kopf Der Mund Der Rücken	Die Backe Die Haare (plural) Die Hand Die Nase Die Schulter Die Zehe Die Zunge	Das Auge Das Bein Das Kinn Das Knie Das Ohr

Section Three: Numbers

Language:

Null Eins Zwei Drei Vier Fünf Sechs Sieben Acht Neun Zehn	 Elf Zwölf Dreizehn Vierzehn Fünfzehn Sechzehn Siebzehn Achtzehn Neunzehn Zwanzig	Die Zahlen

Section Four: General Introductions – more advanced 1

Language:

1. How old are you? Wie alt bist du? Ich bin neun/ zehn Jahre alt 3. (Welcher Nationalität bist du?) Welche Nationalität hast du? Welche Staatsangehörigkeit hast du? Ich bin Irin (fem.) Ich bin Ire (masc.) Die Iren	2. Where do you come from? Woher kommst du? Ich komme aus Irland/ Deutschland/ Frankreich/ den USA/ Österreich/ der Schweiz

Section Five: General Introductions – more advanced 2

Language:

1. Polite form of *you* Wie heißen Sie? (register) Ich heiße _____ Wie alt sind Sie? Ich bin ___ Jahre alt. Woher kommen Sie? Ich komme aus _____	2. *He/ his* Wie heißt er? Er heißt _____ Wie alt ist er? Er ist ___ Jahre alt. Woher kommt er? Er kommt aus _____	3. *She/ her* Wie heißt sie? Sie heißt _____ Wie alt ist sie? Sie ist ___ Jahre alt. Woher kommt sie? Sie kommt aus _____

Section Six: Finding out what things are called

Language:

Was ist das? Das ist ein(e)_____	Wie heißt das auf Deutsch? Das heißt (no article) _____ Das ist ein(e) _____

Ein/ Der Bleistift Ein/ Der Computer Ein/ Der Kuli Ein/ Der Schrank Ein/ Der Stuhl Ein/ Der Tisch Ein/ Der (Wasser)Hahn	Eine/ Die Kreide Eine/ Die Lehrerin Eine/ Die Schere Eine/ Die Schublade Eine/ Die Spüle Eine/ Die Tafel Eine/ Die Tasche Eine/ Die Tür	Ein/ Das Bild Ein/ Das Buch Ein/ Das Fenster Ein/ Das Glas Ein/ Das Heft Ein/ Das Lineal Ein/ Das Regal Ein/ Das Spülbecken

Section Seven: Family

Language:

1. Siblings Hast du/ Haben Sie Geschwister? Ja, ich habe… Eine Schwester/ zwei Schwestern Keine Schwester Einen Bruder/ zwei Brüder Keinen Bruder Nein, ich bin Einzelkind Ich habe keine Geschwister	2. Name and age revised Wie heißt er/sie? Er/ sie heißt _____ Wie alt ist er/sie? Er/ sie ist _____ Jahre alt 3. Younger/Older Ich bin der/die Jüngste/Älteste Ich bin der/die Zweitjüngste/Zweitälteste …ist jünger/älter als ich

Der Vater Der Sohn Der Bruder Der Großvater Der Opa Der Enkel Der Neffe Der Vetter Der Onkel Der Cousin Der Mann Der Freund Der Junge	Die Mutter Die Tochter Die Schwester Die Großmutter Die Oma Die Enkelin Die Nichte Die Kusine Die Tante Die Cousine Die Frau Die Freundin	Die Eltern (pl.) Das Mädchen

Section Eight: Weather

Language:

Das Wetter Wie ist das Wetter (heute)? Was für (ein) Wetter haben wir (heute)? Wieviel Grad haben wir heute? Die Wettervorhersage Das Wetter ist wechselhaft	Der Regen Der Schnee Der Hagel Der Sturm Der Wind Der Nebel Der Frost Die Sonne Die Wolke Das Gewitter Der/ Ein Blitz Der Donner Der Regenschirm Der Regenbogen Der Regentropfen Der Niederschlag Der Schneefall Die Luft Die Nässe Die Windstille Das Eis Das Regenwetter	Es regnet Es schneit Es hagelt Es stürmt Es ist stürmisch Es ist windig Es ist neblig Es ist frostig Es ist sonnig Die Sonne scheint Es ist wolkig/bewölkt Es gibt ein Gewitter Es ist warm/heiß/kalt Es ist nass/trocken Es ist schön/schlecht Es ist schwül Es friert

Use of adjectives with weather: Es ist ein kalt**er** nass**er** … Tag.

Section Nine: Family 2 – Describing yourself and your family

Language:

Mein Vater ist… Dein Vater ist… Ich bin…	Meine Mutter ist… Deine Mutter ist… du bist…	Mein Hund ist … Dein Hund ist… er/sie ist…

Groß Klein Dick Schlank Dünn Alt Jung Blond Dunkelhaarig Stark Schwach Gutaussehend	Nett Freundlich Komisch Schüchtern Höflich Unhöflich Großzügig Fröhlich Schlecht gelaunt

Section Ten: Describing yourself 2

Language:

Ich habe… blaue/braune/grüne Augen braune/blonde/rote/schwarze Haare glatte/lockige Haare kurze/lange Haare Sommersprossen Mein Haar ist/Meine Haare sind… Meine Augen sind…	Er/sie hat… einen Schnurrbart einen Bart er/sie trägt… eine Brille

Section Eleven: School (including clothes)

Language:

Ich lerne… Englisch Irisch Mathe(matik) Erdkunde Geschichte Musik Religion Sport Naturwissenschaft Deutsch Bildende Kunst (visual art) Sozialkunde/Ethik [SPHE (social, personal and health education)]	lernen lesen schreiben etwas fragen spielen rechnen bestehen durchfallen einfach schwerig langweilig interessant	Montags machen wir/mache ich … Am Montag machen wir/mache ich… Mein Lieblingsfach ist… Meine Lieblingsfächer sind…und… Ich mag… Ich hasse…

Der Lehrer Der Schüler Der Füller Der Student Der Kugelschreiber Der Bleistift Der Radiergummi Der Buntstift Der Schulhof Der Stundenplan Der Direktor	Die Karte Die Tafel Die Kreide Die Schülerin Die Studentin Die Schultasche Die Tasche Die Pause Die Klingel Die Garderobe Die Prüfung Die Lehrerin Die (Unterrichts)stunde Die Schule Die Direktorin	Das Fach Das Heft Das Etui Das Lineal Das Schuljahr Das Klassenzimmer

Mein Lehrer heißt Herr…	Ich bin in der vierten/fünften/sechsten Klasse

Meine Lehrerin heißt Frau… Die Schule fängt um 9 Uhr an Die Schule beginnt um 9 Uhr Die Schule ist um 2:30/halb 3 zu Ende Die Schule hört um 3 Uhr auf	Es gibt XX Mädchen und XX Jungen in unserer Klasse Meine/Unsere Schule heißt…

Wir müssen/Ich muss eine (Schul)Uniform tragen Wir tragen/Ich trage keine (Schul)Uniform Wir tragen/Ich trage… Was trägst du jetzt?		
*Roten/blauen... Einen Rock Einen Pullover/Pulli Einen Blazer Einen Schal Einen (Regen)Mantel Einen Hut	*Rote/blaue... Eine Bluse Eine Hose Eine Krawatte Eine Jacke Shorts (pl.) Jeans (pl.) Socken (pl.) Handschuhe (pl.) Schuhe (pl.) Turnschuhe (pl.)	*Rotes/blaues… Ein T-Shirt Ein Hemd Ein Kleid
* Adjective occurs between the article and the noun e.g. Ich trage eine weiße Bluse		

Other clothes include:

Der Anorak Der Badeanzug Der Gürtel Der Knopf Der Schal Der Schnurrsenkel Der Schuh Der (Regen)Schirm Der Stiefel (sing.) Der Trainingsanzug	Die Jeans (pl.) Die Mütze Die Sonnenbrille (sing.)	

Section Twelve: Pets

Language:

Hast du Tiere (zu Hause)? Hast du Haustiere? Ja, ich habe einen Hund Ja, ich habe eine Katze Ja, ich habe ein Pferd Nein, ich habe keine (Haus)Tiere.	Ich habe… einen Hund einen Fisch einen Wellensittich einen Hamster einen Frosch einen Igel

Wie heißt er/sie/es? Wie alt ist er/sie/es? Welche Farbe hat dein Tier? (er/sie/es **ist**) Welche Farbe hat er/sie/es? (er/sie/es **ist**) Seit wie lange/wann hast du ihn/sie/es?	eine Katze eine Schlange eine Schildkröte eine Maus eine Schnecke eine Spinne ein Kaninchen ein Pferd ein Meerschweinchen

Section Thirteen: Pastimes and Hobbies

Language:

Was machst du in deiner Freizeit? Was sind deine Hobbys? Was machst du gern? Was machst du abends? Hast du Hobbys? Was für ein Hobby hast du? Welche Hobbys hast du? Wohin gehst du gern? Ich treibe gern Sport Ich mag Sport Ich spiele Fußball Ich spiele Basketball Ich spiele Volleyball Ich spiele Tennis Ich spiele Bowling Ich spiele Rugby Ich spiele Hurling/Camogie Ich bin Mitglied eines Fußballklubs Ich bin in der Schulmannschaft Ich spiele gern Schach Ich spiele gern Karten Ich spiele gern Dame Am liebsten spiele ich …(Tennis) …(Tennis) spiele ich auch sehr gern Ich spiele lieber mit dem Computer	Ich höre gern Musik Ich lese gern Ich sehe gern fern Ich gehe gern kegeln Ich gehe gern spazieren Ich fahre gern Rad Ich singe gern Ich schwimme gern Ich laufe gern Ich tanze gern Ich bastele gern Ich male gern Ich koche gern Ich sammle/tausche Briefmarken Ich reite gern Spielst du Instrumente? Ich spiele Gitarre Ich spiele Klavier Ich spiele Geige Ich spiele Blechflöte (tin whistle) Ich spiele Blockflöte Ich kann (nicht) … spielen

MUSIK Was für Musik magst du? Ich mag… Tanzmusik klassische Musik Instrumentalmusik	 Was ist deine Lieblingsgruppe? **Meine Lieblingsgruppe ist…** Ich finde sie toll/fantastisch/prima Wer ist dein(e) Lieblingssänger(in)?

Volksmusik Popmusik Rockmusik alternative Musik Blues-Musik Jazz Musik aus den Charts	Mein(e) Lieblingssänger(in) ist… Er/sie ist toll/fantastisch/prima Manchmal gehe ich zu einem Konzert

LESEN	FERNSEHEN/KINO
Was für Bücher liest du am liebsten? Was für Bücher magst du? Ich mag/ Ich hasse…… Horrorbücher Krimis romantische Bücher Romane Biographien Autobiographien Reisebücher Zeitschriften Zeitungen Comics Wer ist dein(e) Lieblingsschriftsteller(in)? Ich mag…	Wer ist dein(e) Lieblingsschauspieler(in)? Meine Lieblingssendung ist… Ich mag… Ich hasse… Ich gehe gern ins Kino Ich mag… Serien Seifenoper Filme Cartoons Musiksendungen Talk-Shows Der Kanal Mein Lieblingskanal ist…

Section Fourteen: Numbers 20+

Language:

zwanzig dreißig vierzig fünfzig sechzig siebzig achtzig neunzig hundert tausend	Einundzwanzig Zweiundzwanzig Dreiundzwanzig … einunddreißig zweiunddreißig dreiunddreißig …

Section Fifteen: Time

Language:

Wie spät ist es bitte? Wieviel Uhr ist es bitte? Es ist ein/zwei/drei… Uhr morgens/abends Es ist Viertel nach eins/zwei… Es ist Viertel vor eins/zwei… Es ist halb eins/zwei Es ist fünf/sechs (Minuten) nach neun Es ist fünf/sechs (Minuten) vor neun Es ist Mittag/Mitternacht Ein Uhr zwanzig Dreizehn Uhr vierzehn (13:14) …	Wann stehst du morgens auf? Ich stehe morgens um 8 Uhr auf Am Wochenende stehe ich um 10 Uhr auf. Wann gehst du ins Bett? Um 8 Uhr abends gehe ich ins Bett. Ich verlasse mein Haus um Viertel vor neun. Ich erreiche die Schule um fünf vor neun. Wann hast du deine Pause/Mittagspause? Wann machst du deine Hausaufgaben? …

Section Sixteen: Birthdays and date

Language:

1st	erste	11th	elfte	21st	einundzwanzigste
2nd	zweite	12th	zwölfte	22nd	zweiundzwanzigste
3rd	dritte	13th	dreizehnte	23rd	dreiundzwanzigste
4th	vierte	14th	vierzehnte	24th	vierundzwanzigste
5th	fünfte	15th	fünfzehnte	25th	fünfundzwanzigste
6th	sechste	16th	sechzehnte	26th	sechsundzwanzigste
7th	siebte	17th	siebzehnte	27th	siebenundzwanzigste
8th	achte	18th	achtzehnte	28th	achtundzwanzigste
9th	neunte	19th	neunzehnte	29th	neunundzwanzigste
10th	zehnte	20th	zwanzigste	30th	dreißigste
				31st	einunddreißigste

Januar Februar März April Mai Juni Juli August September Oktober November Dezember	Wann hast du Geburtstag? Mein Geburtstag ist am neunundzwanzigsten März Mein Geburtstag ist im März Wann hat deine Schwester/dein Bruder Geburtstag? Er/sie hat am … Geburtstag Am neunzehnten Juni gehen wir in Urlaub fahren wir in Urlaub fliegen wir in Urlaub (auf Urlaub – mainly used in Austria)	den 29. Januar den 13. März

Section Seventeen: Food and shopping

Language:

Was isst du gern? Magst du…(Salat)? Ich mag/esse gern… Ich mag … nicht Ich hasse… Was magst du am liebsten? Ich mag … am liebsten Ich mag … sehr gern	Der Salat Der Fisch Der Kuchen Der Hamburger Der Käse Der Schinken Der Spinat Der Weißkohl Der Blumenkohl Der Rosenkohl Der Pfirsich(e) Der Apfel	Die Pommes Frites Die Pizza Die Butter Die Spaghetti Die Tomate(n) Die Banane(n) Die Traube(n) Die Wurst Die Kartoffel Die Zwiebel(n) Die Erbse(n) Die Himbeere(n) Die Erdberre(n) Die Melone(n) Die Kirsche(n) Die Aprikose(n) Die Pflaume(n) Die Birne(n) Die Apfelsine(n) Die Zitrone	Das Obst Das Hähnchen Das Brot Das Gemüse Das Fleisch

Ins Geschäft Kann ich Ihnen helfen? Das ist teuer Das ist billig	Was kostet das? Das kostet …… Euro Danke schön

Ich möchte Ich will	einen Karton einen Becher eine Packung eine Dose eine Tafel ein Liter ein Kilo ein Pfund ein Glas ein Stück ein Paket	Orangensaft Margarine Milch Bohnen Schokolade Milch Kartoffeln Butter Honig Kuchen Chips	bitte

Section Eighteen: Fillers

The Alphabet – Das Alphabet

a – ah, apple	h – ha, haha	o – oh	v - fow
b - bay	i - eeh	p - pay	w - vay
c - tsay	j - yacht	q – koo, cuckoo	x – ixx, sticks
d - day	k – kah, cat	r – err, air	y – oop-see-lohn, üpsilon
e – ay, day	l - ell	s - ess	z - zett
f - eff	m - emm	t – tae (Ir.), tay	
g - gay	n - enn	u – ooh, poo	

Days of the week

Montag	Freitag
Dienstag	Samstag
Mittwoch	Sonntag
Donnerstag	

Months of the year

Januar	Juli	Was findet im Dezember statt?
Februar	August	Wann fängt die Schule an?
März	September	Wann gehen/fliegen/fahren wir in Urlaub?
April	Oktober	
Mai	November	Diesen Monat…
Juni	Dezember	Letzten Monat…
		Am 10. dieses Monats…

Colours - Die Farbe(n)

Orange	Rot
Rosa	Schwarz
Violett / lila	Braun
Blau	Grau
Gelb	Weiß
Grün	

Shapes

der Kreis	die Kugel	das Viereck
der Würfel		das Dreieck
der Kegel		das Rechteck
der Zylinder		das Oval
		das Achteck

Seasons - Die Jahreszeit (en)

Der Frühling Der Sommer Der Herbst Der Winter

Section Nineteen: Classroom Language:

Roll-call	**Praise**	**General**
Hier Nicht hier Ja Nein	Gut Sehr gut Fantastisch Wunderbar Prima toll	Darf ich zur Toilette gehen bitte? Darf ich bitte auf die Toilette gehen? Darf ich bitte aufs Klo gehen? Darf ich meinen Bleistift spitzen? Könnten Sie bitte mal herkommen? Könnten Sie das bitte erklären?

Understanding	**Questions**	**Games**
Ich verstehe das nicht. Ich kapiere das nicht Ich habe keine Ahnung Wiederholen Sie (das), bitte Nochmal/ Noch einmal bitte Wie bitte? Jetzt verstehe ich (das)	Ich habe eine Frage Was machen wir heute? Welche Seite? Wo sind wir? (Auf Seite 2) Was bedeutet 'Freund' auf Englisch Wie sagt man 'see you soon' auf Deutsch? Was ist los? Wer ist das? Bist du sicher? Verstehst du? Wo ist…? Alles klar?	Würfle! Ich bin dran Du bist dran Wer ist dran? Kein Betrügen! Ich habe gewonnen Ich habe verloren Welche Nummer hast du bekommen? Wie viele Karten hast du?

Excuses	**Extras**	**Opinion**
Ich habe meine Hausaufgaben vergessen Ich habe nichts auf. Mein Hund hat die Hausaufgaben gefressen! Was war in Deutsch auf? Haben wir heute Hausaufgaben?	Vielleicht Später Jetzt Danke (schön/sehr) Bitte (schön/sehr) Kein Problem Das macht nichts Nichts zu danken Es tut mir leid Steh auf! Setz dich! Mach die Tür zu! Öffne die Tür! Mach die Tür auf! Halt Sei ruhig!	Das ist einfach/ leicht Das ist schwer Deutsch macht Spaß Spaß muss sein!

Appendix E Raw linguistic content version 1.1

Addition of linkage, integration and language awareness within Phase I.

Section One: General Introductions

Language:

<table>
<tr><td>1. Greetings

Hallo!
Guten Morgen
Guten Tag
Guten Abend
Gute Nacht
Schlaf gut</td><td>

Auf Wiedersehen
Tschüs
Bis bald
Bis morgen/Donnerstag…</td></tr>
<tr><td>2. How are you?

Wie geht’s?
Gut, danke!
Sehr gut, danke!
Nicht so gut
Schlecht
Mir geht’s gut/sehr gut…</td><td>3. What is your name?

Wie heißt du?
Ich heiße _____</td></tr>
<tr><td>Integration
• Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >téamaí: Mé féin
• English> oral language, reading and writing>competence and confidence in using language and receptiveness to language
• History>Myself and my family/local studies>Myself
• SPHE>Myself>Self-identity
• SPHE>Myself and Others>Relating to others

Linkage
• Section Four: General Introductions – more advanced 1
• Section Five: General Introductions – more advanced 2
• Section Six: Finding out what things are called
• Section Seven: Family
• Section Twelve: Pets</td><td>Language Awareness
• Word order for naming yourself in Irish differs from German and English – [] is ainm dom, ich heiße [], my name is [].
• Examine ‘hello’ expressions in German, English and Irish. Note ‘dia dhuit’ and its response(s).
• Examine dialects in Irish – three ways of asking ‘how are you?’ - Chaoi a bhfuil tú? Conas atá tú? Cad é mar atá tú?
• Examine contractions in English (e.g. how’re/how are), Irish (e.g. Táim/Tá mé) and German (e.g. geht’s/geht es)
• Examine how questions are asked in each of the languages</td></tr>
</table>

Section Two: Parts of the body

Language:

Der Bauch Der Daumen Der Finger Der Fuß Der Hals Der Kopf Der Mund Der Rücken	Die Backe Die Haare (plural) Die Hand Die Nase Die Schulter Die Zehe Die Zunge	Das Auge Das Bein Das Kinn Das Knie Das Ohr

Integration	Language Awareness
• Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >téamaí: Mé féin • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • SPHE>Taking care of my body>Knowing about my body • Science>Living things>Myself/Human life Linkage • Section Twelve: Pets	• Examine words which look/sound like the English equivalent (cognates) (e.g. Finger, Fuß, Nase, Haare, Schulter, Hand, Knie). • Examine language families as the reason why there are many English/German similarities and few Irish/German ones • Examine English words which contain German ones/have their root in German – Rucksack/back pack from the German 'Rücken' • Describe nouns – what is a noun? • Examine how all nouns in German receive a capital letter. Examine the use of capital letters in English and Irish. • Examine the use of gender with nouns in German (masculine, feminine and neuter) and Irish (masculine and feminine)

Section Three: Numbers

Language:

Null Eins Zwei Drei Vier Fünf Sechs Sieben Acht Neun Zehn	Elf Zwölf Dreizehn Vierzehn Fünfzehn Sechzehn Siebzehn Achtzehn Neunzehn Zwanzig	Die Zahlen

Integration	Language Awareness
• Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht > • English> oral language, reading and writing>competence and confidence in	• Examine words which look/sound like the English equivalent and mean the same thing (cognates) (e.g. Telefonnummer) or

using language and receptiveness to language • Mathematics>Number>Counting and numeration and comparing and ordering • Mathematics>Measures>Money Linkage • Section Four: General Introductions – more advanced 1 • Section Five: General Introductions – more advanced 2 • Section Seven: Family • Section Fourteen: Numbers 20+ • Section Fifteen: Time • Section Sixteen: Birthdays and date • Section Seventeen: Food (and shopping)	Irish equivalent (e.g. acht/a hocht, Telefon/fón/theileafóin). • Examine words which look/sound like an English word, but which do not mean the same/similar things (false friends) (e.g. drei/dry, vier/fear, elf/elf, zwölf/wolf) • Examine language families as the reason why there are many English/German similarities and few Irish/German ones • Examine how numbering schemes work in English, German and Irish – e.g. addition of 'teen', 'zehn' and 'déag' • Examine word order, Irish is a verb initial language (VSO: Verb Subject Object) and English is an SVO language. German is also an SVO language, with SOV subordinate clauses. • Examine compound words in German – e.g. drei|zehn, vier|zehn

Section Four: General Introductions – more advanced 1

Language:

1. How old are you? Wie alt bist du? Ich bin neun/ zehn Jahre alt 3. (Welcher Nationalität bist du?) Welche Nationalität hast du? Welche Staatsangehörigkeit hast du? Ich bin Irin (fem.) Ich bin Ire (masc.) Die Iren	2. Where do you come from? Woher kommst du? Ich komme aus Irland/ Deutschland/ Frankreich/ den USA/ Österreich/ der Schweiz
Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >téamaí: Mé féin • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • History>Myself and my family/local studies>Myself • SPHE>Myself>Self-identity • SPHE>Myself and Others>Relating to others • SPHE>Myself and the wider world>Developing citizenship • Mathematics>Number>Counting and numeration; Comparing and ordering	Language Awareness • Examine words which look/sound like the English equivalent and mean the same thing (cognates) (e.g. Nationalität, Irland, USA, kommen) or Irish equivalent (e.g. náisiúntacht, Fhrainc). • Examine the use of sein and haben in comparison to English – e.g. what nationality **are** you, ich habe Hunger/I have hunger/I am hungry • Examine masculine and feminine nouns and inflection in German and English (e.g. actor/actress)

• Geography>Human Environments>Living in the local community; People and places in other areas; People and other lands Linkage • Section One: General Introductions • Section Three: Numbers 1-20 • Section Five: General Introductions – more advanced 2 • Section Six: Finding out what things are called • Section Seven: Family • Section Twelve: Pets	

Section Five: General Introductions – more advanced 2

Language:

1. Polite form of *you*	2. *He/ his*	3. *She/ her*
Wie heißen Sie? (register) Ich heiße _____ Wie alt sind Sie? Ich bin ___ Jahre alt. Woher kommen Sie? Ich komme aus _____	Wie heißt er? Er heißt _____ Wie alt ist er? Er ist ___ Jahre alt. Woher kommt er? Er kommt aus _____	Wie heißt sie? Sie heißt _____ Wie alt ist sie? Sie ist ___ Jahre alt. Woher kommt sie? Sie kommt aus _____

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >téamaí: Mé féin • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • History>Myself and my family/local studies>Myself • SPHE>Myself>Self-identity • SPHE>Myself and Others>Relating to others • SPHE>Myself and the wider world>Developing citizenship • Mathematics>Number>Counting and numeration; Comparing and ordering • Geography>Human Environments>Living in the local community; People and places in other areas; People and other lands Linkage • Section One: General Introductions –	Language Awareness • Examine the use of the polite form in German. • Examine the importance of capital letters for nouns and pronouns (polite form) in German. • Examine the use of personal pronouns across the three languages.

• Section Three: Numbers 1-20 • Section Four: General Introductions – more advanced 1 • Section Six: Finding out what things are called • Section Seven: Family • Section Twelve: Pets	

Section Six: Finding out what things are called

Language:

Was ist das? Das ist ein(e)_____	Wie heißt das auf Deutsch? Das heißt (no article) _____ Das ist ein(e) _____

Ein/ Der Bleistift Ein/ Der Computer Ein/ Der Kuli Ein/ Der Schrank Ein/ Der Stuhl Ein/ Der Tisch Ein/ Der (Wasser)Hahn	Eine/ Die Kreide Eine/ Die Lehrerin Eine/ Die Schere Eine/ Die Schublade Eine/ Die Spüle Eine/ Die Tafel Eine/ Die Tasche Eine/ Die Tür	Ein/ Das Bild Ein/ Das Buch Ein/ Das Fenster Ein/ Das Glas Ein/ Das Heft Ein/ Das Lineal Ein/ Das Regal Ein/ Das Spülbecken

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >téamaí: An Scoil • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • History>Myself and my family/local studies>My school • SPHE>Myself and the wider world>My school community • Geography>Human Environments>Living in the local community (School); Linkage (Focus: heiβen, relevant vocabulary) • Section One: General Introductions • Section Four: General Introductions – more advanced 1 • Section Five: General Introductions – more advanced 2 • Section Seven: Family • Section Eleven: School • Section Twelve: Pets	Language Awareness • Examine words which look/sound like the English equivalent (cognates) (e.g. Computer, Stuhl, Wasser, Glas). • Examine how all nouns in German receive a capital letter. Examine the use of capital letters in English and Irish. • Examine the use of gender with nouns in German (masculine, feminine and neuter) and Irish (masculine and feminine)

Section Seven: Family

Language:

1. Siblings Hast du/ Haben Sie Geschwister? Ja, ich habe… Eine Schwester/ zwei Schwestern Keine Schwester Einen Bruder/ zwei Brüder Keinen Bruder Nein, ich bin Einzelkind Ich habe keine Geschwister	2. Name and age revised Wie heißt er/sie? Er/ sie heißt _____ Wie alt ist er/sie? Er/ sie ist _____ Jahre alt 3. Younger/Older Ich bin der/die Jüngste/Älteste Ich bin der/die Zweitjüngste/Zweitälteste …ist jünger/älter als ich

Der Vater	Die Mutter	Die Eltern (pl.)
Der Sohn	Die Tochter	
Der Bruder	Die Schwester	
Der Großvater	Die Großmutter	
Der Opa	Die Oma	
Der Enkel	Die Enkelin	
Der Neffe	Die Nichte	
Der Vetter	Die Kusine	
Der Onkel	Die Tante	
Der Cousin	Die Cousine	
Der Mann	Die Frau	
Der Freund	Die Freundin	
Der Junge		Das Mädchen

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >téamaí: Mé féin; Sa bhaile • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • History>Myself and my family/local studies>Myself • SPHE>Myself>Self-identity • SPHE>Myself and Others>Relating to others • SPHE>Myself and the wider world>Developing citizenship • Mathematics>Number>Counting and numeration; Comparing and ordering • Geography>Human Environments>Living in the local community; People and places in other areas; People and other lands Linkage • Section One: General Introductions • Section Four: General Introductions – more advanced 1	Language Awareness • Examine words which look/sound like the English equivalent (cognates) (e.g. Neffe, Onkel, Cousin, Mann, Freund, Kusine, Counsine, Freundin, Jüngste, Bruder) or the Irish equivalent (e.g. An t-uncail, an mháthair). • Examine how plural nouns are formed for Schwester(n) and Bruder(¨), and how they are formed in English and Irish. • Examine how there is a male and female form of some nouns in German – Freund, Freundin (like English actor, actress; waiter, waitress) • Examine how all nouns in German receive a capital letter. Examine the use of capital letters in English and Irish. • Examine the use of gender with nouns in German (masculine, feminine and neuter) and Irish (masculine and feminine) • Examine counting people in Irish (vs. things) • Examine compound words – Einzel|kind, Zweit|jüngste

• Section Five: General Introductions – more advanced 2 • Section Nine: Family II and Describing Yourself I • Section Ten: Describing Yourself II • Section Twelve: Pets	• Describe pronouns, what is a pronoun? • Examine personal pronouns across the three languages • Describe an adjective, what is an adjective? • Examine the comparative and the superlative in each language (e.g. old, older, oldest / sean, níos sine, is sine / alt, alter, Älteste

Section Eight: Weather

Language:

Das Wetter	Der Regen	Es regnet
	Der Schnee	Es schneit
Wie ist das Wetter (heute)?	Der Hagel	Es hagelt
Was für (ein) Wetter haben wir (heute)?	Der Sturm	Es stürmt
		Es ist stürmisch
Wieviel Grad haben wir heute?	Der Wind	Es ist windig
	Der Nebel	Es ist neblig
Die Wettervorhersage	Der Frost	Es ist frostig
	Die Sonne	Es ist sonnig
Das Wetter ist wechselhaft		Die Sonne scheint
	Die Wolke	Es ist wolkig/bewölkt
	Das Gewitter	Es gibt ein Gewitter
	Der/ Ein Blitz	Es ist warm/heiß/kalt
	Der Donner	Es ist nass/trocken
	Der Regenschirm	Es ist schön/schlecht
	Der Regenbogen	Es ist schwül
	Der Regentropfen	Es friert
	Der Niederschlag	
	Der Schneefall	
	Die Luft	
	Die Nässe	
	Die Windstille	
	Das Eis	
	Das Regenwetter	

Use of adjectives with weather: Es ist ein kalt**er** nass**er** … Tag.

Integration	Language Awareness
• Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >téamaí: An Aimsir • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • SPHE>Myself and the wider world>Environmental care • Mathematics>Number> Comparing and ordering • Geography>Natural Environments>Weather, climate and atmosphere	• Examine words which look/sound like the English equivalent (cognates) (e.g. Wind, Wetter, Frost, Sonne, Regen, Schnee, Eis). • Examine how all nouns in German receive a capital letter. Examine the use of capital letters in English and Irish. • Examine the use of gender with nouns in German (masculine, feminine and neuter) and Irish (masculine and feminine) • Examine compound words – Schnee|fall, Wind|stille, Regen|wetter, Regen|bogen • Examine the comparative and the superlative in each language (e.g. cold, colder, coldest)

• Science>Environmental awareness and care>Environmental awareness; Caring for the environment Linkage • Section Six: Finding out what things are called • Section Thirteen: Pastimes and Hobbies (when the sun shines, when it's raining…) • Fillers: Months of the year, Seasons	• Describe an adjective, what is an adjective? • Examine the use of adjectives across the three languages and the position of adjectives in a sentence (e.g.it is a *cold* day, es ist ein *kalter* Tag, is lá *fuar* é).

Section Nine: Family 2 – Describing yourself and your family

Language:

Mein Vater ist… Dein Vater ist… Ich bin…	Meine Mutter ist… Deine Mutter ist… du bist…	Mein Hund ist … Dein Hund ist… er/sie ist…

Groß Klein Dick Schlank Dünn Alt Jung Blond Dunkelhaarig Stark Schwach Gutaussehend	Nett Freundlich Komisch Schüchtern Höflich Unhöflich Großzügig Fröhlich Schlecht gelaunt

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >téamaí: Mé féin; Sa bhaile • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • History>Myself and my family/local studies>Myself • SPHE>Myself>Self-identity • SPHE>Myself and Others>Relating to others Linkage • Section One: General Introductions • Section Four: General Introductions – more advanced 1 • Section Five: General Introductions – more advanced 2 • Section Seven: Family • Section Ten: Describing Yourself II • Section Twelve: Pets • Fillers: Colour	Language Awareness • Examine words which look/sound like the English equivalent (cognates) (e.g. Hund (hound), Mutter, Vater, jung, blond) or the Irish equivalent (e.g. an mháthair). • Describe pronouns, what is a pronoun? • Examine personal pronouns across the three languages and how they affect the noun – e.g. in Irish, mo+h, do+h. • Describe yourself using the three languages.

Section Ten: Describing yourself 2

Language:

Ich habe… blaue/braune/grüne Augen braune/blonde/rote/schwarze Haare glatte/lockige Haare kurze/lange Haare Sommersprossen Mein Haar ist/Meine Haare sind… Meine Augen sind…	Er/sie hat… einen Schnurrbart einen Bart er/sie trägt… eine Brille

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >téamaí: Mé féin; Sa bhaile • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • History>Myself and my family/local studies>Myself • SPHE>Myself>Self-identity • SPHE>Myself and Others>Relating to others Linkage • Section One: General Introductions • Section Four: General Introductions – more advanced 1 • Section Five: General Introductions – more advanced 2 • Section Seven: Family I • Section Nine: Family II and Describing Yourself I • Section Ten: Describing Yourself I • Section Twelve: Pets • Fillers: Colour	Language Awareness • Examine words which look/sound like the English equivalent (cognates) (e.g. blau, braun, blond, grün). • Examine compound words: Sommer|sprossen • Describe an adjective, what is an adjective? • Examine the use of adjectives across the three languages and the position of adjectives in a sentence (e.g. he has black hair, er hat *schwarze* Haare, tá gruaig *dhubh* air).

Section Eleven: School (including clothes)

Language:

Ich lerne…	lernen lesen	Montags machen wir/mache ich …
Englisch	schreiben	Am Montag machen wir/mache ich…
Irisch	etwas fragen	
Mathe(matik)	spielen	Mein Lieblingsfach ist…
Erdkunde	rechnen	Meine Lieblingsfächer sind…und…
Geschichte	bestehen	
Musik	durchfallen	Ich mag…
Religion		Ich hasse…

Sport Naturwissenschaft Deutsch Bildende Kunst (visual art) Sozialkunde/Ethik [SPHE (social, personal and health education)]	einfach schwerig langweilig interessant	

Der Lehrer Der Schüler Der Füller Der Student Der Kugelschreiber Der Bleistift Der Radiergummi Der Buntstift Der Schulhof Der Stundenplan Der Direktor	Die Karte Die Tafel Die Kreide Die Schülerin Die Studentin Die Schultasche Die Tasche Die Pause Die Klingel Die Garderobe Die Prüfung Die Lehrerin Die (Unterrichts)stunde Die Schule Die Direktorin	Das Fach Das Heft Das Etui Das Lineal Das Schuljahr Das Klassenzimmer

Mein Lehrer heißt Herr… Meine Lehrerin heißt Frau… Die Schule fängt um 9 Uhr an Die Schule beginnt um 9 Uhr Die Schule ist um 2:30/halb 3 zu Ende Die Schule hört um 3 Uhr auf	Ich bin in der vierten/fünften/sechsten Klasse Es gibt XX Mädchen und XX Jungen in unserer Klasse Meine/Unsere Schule heißt…

Wir müssen/Ich muss eine (Schul)Uniform tragen Wir tragen/Ich trage keine (Schul)Uniform Wir tragen/Ich trage… Was trägst du jetzt?		
*Roten/blauen... Einen Rock Einen Pullover/Pulli Einen Blazer Einen Schal Einen (Regen)Mantel Einen Hut	*Rote/blaue... Eine Bluse Eine Hose Eine Krawatte Eine Jacke Shorts (pl.) Jeans (pl.) Socken (pl.) Handschuhe (pl.) Schuhe (pl.) Turnschuhe (pl.)	*Rotes/blaues… Ein T-Shirt Ein Hemd Ein Kleid
* Adjective occurs between the article and the noun e.g. Ich trage eine weiße Bluse		

Other clothes include:

Der Anorak Der Badeanzug Der Gürtel Der Knopf Der Schal Der Schnurrsenkel Der Schuh Der (Regen)Schirm Der Stiefel (sing.) Der Trainingsanzug	Die Jeans (pl.) Die Mütze Die Sonnenbrille (sing.)	

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >Téamaí: Mé féin; An Scoil; Éadaí • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • History>Myself and my family/local studies>My school • History>Change and continuity>Clothes • SPHE>Myself and the wider world>My school community • Geography>Human Environments>Living in the local community (School) Linkage • Section One: General Introductions • Section Three: Numbers 1-20 • Section Four: General Introductions – more advanced 1 • Section Five: General Introductions – more advanced 2 • Section Six: Finding out what things are called • Section Nine: Family II and Describing Yourself I • Section Ten: Describing Yourself II • Section Twelve: Pets • Section Fourteen: Numbers 20+ • Section Fifteen: Time • Section Sixteen: Birthdays and Date • Section Nineteen: Classroom Language • Fillers: Colour	Language Awareness • Examine words which look/sound like the English equivalent (cognates) (e.g.Anorak, Schal, Schuh, Jeans, Socken, Jacke, T-shirt, Englisch, Irisch, Mathe(matik), Musik, Religion, Sport, lernen, interresant, Montag, Student, Direktor, Karte, Pause, beginnen) or Irish (e.g. Reiligiún, Mata, Gearmáinis. ag staidéar, peann, cárta, bord, sciorta, scairf, hata, blús, stocaí, an t-anorac). • Examine words which look like English words, but which don't relate in meaning (false friends) (e.g. Rock, fangen, Art, Note) • Examine separable verbs in German: aufhoren, anfangen • Examine how all nouns in German receive a capital letter. Examine the use of capital letters in English and Irish. • Examine the use of gender with nouns in German (masculine, feminine and neuter) and Irish (masculine and feminine) • Examine compound words – Hand|schuhe, Sonnen|brille, Regen|schirm, Schul|tasche • Examine how we talk about thing that we do and don't like (graded): I like, I love, I don't like, I hate • Describe an adjective, what is an adjective? • Examine the use of adjectives across the three languages and the position of adjectives in a sentence (e.g. I wear a *white* blouse, Ich trage eine *weiße* Bluse, Tá blús *bán* á chaitheamh agam).

Section Twelve: Pets

Language:

Hast du Tiere (zu Hause)? Hast du Haustiere?	Ich habe…

Ja, ich habe einen Hund Ja, ich habe eine Katze Ja, ich habe ein Pferd Nein, ich habe keine (Haus)Tiere. Wie heißt er/sie/es? Wie alt ist er/sie/es? Welche Farbe hat dein Tier? (er/sie/es **ist**) Welche Farbe hat er/sie/es? (er/sie/es **ist**) Seit wie lange/wann hast du ihn/sie/es?	einen Hund einen Fisch einen Wellensittich einen Hamster einen Frosch einen Igel eine Katze eine Schlange eine Schildkröte eine Maus eine Schnecke eine Spinne ein Kaninchen ein Pferd ein Meerschweinchen

Integration	Language Awareness
• Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >téamaí: Mé féin; Sa bhaile; Caitheamh Aimsire (peata) • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • History>Myself and my family/local studies>My school • SPHE>Myself and others> Myself and my family; Relating to others • Science>Plants and animals>Variety and characteristics of living things Linkage • Section One: General Introductions • Section Two: Parts of the Body (naming on pets) • Section Four: General Introductions – more advanced 1 • Section Five: General Introductions – more advanced 2 • Section Seven: Family • Section Nine: Family II and Describing Yourself I • Section Ten: Describing Yourself II • Fillers: Colour	• Examine words which look/sound like the English equivalent (cognates) (e.g. Haus, Hund (hound), Katze, Maus, Fisch, Hamster, Frosch) or Irish (e.g. peata, cat, toirtís). • Examine words which look like English words, but which don't relate in meaning (false friends) (e.g. Spinne, Tier) • Examine how all nouns in German receive a capital letter. Examine the use of capital letters in English and Irish. • Examine the use of gender with nouns in German (masculine, feminine and neuter) and Irish (masculine and feminine) • Examine compound words – Haus|tiere, Meer|schwein|chen • Describe an adjective, what is an adjective? • Examine the use of adjectives across the three languages and the position of adjectives in a sentence (e.g. He is *brown*, er ist *braun*, tá sé *donn*).

Section Thirteen: Pastimes and Hobbies

Language:

Was machst du in deiner Freizeit? Was sind deine Hobbys? Was machst du gern? Was machst du abends?	Ich höre gern Musik Ich lese gern Ich sehe gern fern Ich gehe gern kegeln

Hast du Hobbys? Was für ein Hobby hast du? Welche Hobbys hast du? Wohin gehst du gern? Ich treibe gern Sport Ich mag Sport Ich spiele Fußball Ich spiele Basketball Ich spiele Volleyball Ich spiele Tennis Ich spiele Bowling Ich spiele Rugby Ich spiele Hurling/Camogie Ich bin Mitglied eines Fußballklubs Ich bin in der Schulmannschaft Ich spiele gern Schach Ich spiele gern Karten Ich spiele gern Dame Am liebsten spiele ich …(Tennis) …(Tennis) spiele ich auch sehr gern Ich spiele lieber mit dem Computer	Ich gehe gern spazieren Ich fahre gern Rad Ich singe gern Ich schwimme gern Ich laufe gern Ich tanze gern Ich bastele gern Ich male gern Ich koche gern Ich sammle/tausche Briefmarken Ich reite gern Spielst du Instrumente? Ich spiele Gitarre Ich spiele Klavier Ich spiele Geige Ich spiele Blechflöte (tin whistle) Ich spiele Blockflöte Ich kann (nicht) … spielen

MUSIK Was für Musik magst du? Ich mag… Tanzmusik klassische Musik Instrumentalmusik Volksmusik Popmusik Rockmusik alternative Musik Blues-Musik Jazz Musik aus den Charts	 Was ist deine Lieblingsgruppe? **Meine Lieblingsgruppe ist…** Ich finde sie toll/fantastisch/prima Wer ist dein(e) Lieblingssänger(in)? Mein(e) Lieblingssänger(in) ist… Er/sie ist toll/fantastisch/prima Manchmal gehe ich zu einem Konzert

LESEN Was für Bücher liest du am liebsten? Was für Bücher magst du? Ich mag/ Ich hasse…… Horrorbücher Krimis romantische Bücher Romane Biographien	FERNSEHEN/KINO Wer ist dein(e) Lieblingsschauspieler(in)? Meine Lieblingssendung ist… Ich mag… Ich hasse… Ich gehe gern ins Kino Ich mag…

Autobiographien Reisebücher Zeitschriften Zeitungen Comics Wer ist dein(e) Lieblingsschriftsteller(in)? Ich mag…	Serien Seifenoper Filme Cartoons Musiksendungen Talk-Shows Der Kanal Mein Lieblingskanal ist…

Integration	Language Awareness
• Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >téamaí: Mé féin; Caitheamh aimsire; An teilifís • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • History>Myself and my family/local studies>Myself • History>Myself and My Family/Local Studies>Games and pastimes in the past • Geography>Human Environments>Living in the local community>People at work; People at play • SPHE>Myself>Self-identity • SPHE>Myself and Others>Relating to others • P.E.>Athletics; Dance; Games; Outdoor and Adventure activities • Visual Arts>Drawing; Paint and colour; Construction; Fabric and fibre Linkage • Section Eight: Weather (In good and bad weather, I like to…) • Section Eleven: School (I play/do … in school) • Section Seventeen: Food (I like to cook …)	• Examine words which look/sound like the English equivalent (cognates) (e.g. Musik, Hobbys, Sport, schwimmen, Fuβball, Basketball, Volleyball, Tennis, Bowling, Rugby, Hurling, Klub, Karten, Computer, kochen, Instrumente, Musik, Gruppe, klassiche Musik, Instrumentalmusik, Popmusik, Rockmusik, alternative Musik, Blues-Musik, Jazz, Konzert, Buch, Biographien, Autobiographien, Comics, Serien, Filme, Talk-Shows) or Irish (e.g. Spórt, rugbaí, cártaí, giotár, pianó, grúpaí, cartúin, stáisiún teilifíse). • Examine words which look like English words, but which don't relate in meaning (false friends) (e.g. Romane) • Examine how all nouns in German receive a capital letter. Examine the use of capital letters in English and Irish. • Examine the use of gender with nouns in German (masculine, feminine and neuter) and Irish (masculine and feminine) • Examine compound words – Fuβball|klubs, Tanz|music, Instrumental|musik, Lieblings|sending, Lieblings|sänger(in), Lieblings|schauspieler(in) • Examine how we express likes and dislikes (graded) – I like, I love, I don't like, I hate, my favourite…

Section Fourteen: Numbers 20+

Language:

zwanzig dreißig vierzig fünfzig sechzig siebzig achtzig neunzig	Einundzwanzig Zweiundzwanzig Dreiundzwanzig … einunddreißig zweiunddreißig dreiunddreißig

hundert tausend	…

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • Mathematics>Number>Counting and numeration; Comparing and ordering Linkage: • Section Three: Numbers 1-20 • Section Four: General Introductions: More Advanced I • Section Five: General Introductions: More Advanced II • Section Seven: Family I • Section Fifteen: Time • Section Sixteen: Birthdays and Date	Language Awareness • Examine words which look/sound like the Irish equivalent (cognates) (e.g. acht) • Examine words which look like English words, but which don't relate in meaning (false friends) (e.g. drei) • Examine compound words – drei|zig, ein|und|drei|zig • Examine the numbering system across the three languages • Examine how to count people in Irish (vs. things)

Section Fifteen: Time

Language:

Wie spät ist es bitte? Wieviel Uhr ist es bitte? Es ist ein/zwei/drei… Uhr morgens/abends Es ist Viertel nach eins/zwei… Es ist Viertel vor eins/zwei… Es ist halb eins/zwei Es ist fünf/sechs (Minuten) nach neun Es ist fünf/sechs (Minuten) vor neun Es ist Mittag/Mitternacht Ein Uhr zwanzig Dreizehn Uhr vierzehn (13:14) …	Wann stehst du morgens auf? Ich stehe morgens um 8 Uhr auf Am Wochenende stehe ich um 10 Uhr auf. Wann gehst du ins Bett? Um 8 Uhr abends gehe ich ins Bett. Ich verlasse mein Haus um Viertel vor neun. Ich erreiche die Schule um fünf vor neun. Wann hast du deine Pause/Mittagspause? Wann machst du deine Hausaufgaben? …

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >Mé féin; Sa bhaile; An scoil • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • Mathematics>Number>Counting and numeration and comparing and ordering	Language Awareness • Examine words which look/sound like the English equivalent (cognates) (e.g. Morgens, Minuten, Bett, Haus, halb) or the Irish equivalent (e.g. maidin) • Examine words which look like English words, but which don't relate in meaning (false friends) (e.g. Morgen, Pause)

• Mathematics>Measures>Time • History>Myself and my family/local studies>My school • SPHE>Myself and the wider world>My school community • Geography>Human Environments>Living in the local community (School) Linkage • Section Three: Numbers 1-20 • Section Eleven: School (At X time, we…) • Section Fourteen: Numbers 20+	• Examine compound words – drei|zig, Mittags|pause, Haus|aufgaben, Wochen|ende • Examine the numbering system across the three languages • Examine how to count people in Irish (vs. things)

Section Sixteen: Birthdays and date

Language:

1st	erste	11th	elfte	21st	einundzwanzigste
2nd	zweite	12th	zwölfte	22nd	zweiundzwanzigste
3rd	dritte	13th	dreizehnte	23rd	dreiundzwanzigste
4th	vierte	14th	vierzehnte	24th	vierundzwanzigste
5th	fünfte	15th	fünfzehnte	25th	fünfundzwanzigste
6th	sechste	16th	sechzehnte	26th	sechsundzwanzigste
7th	siebte	17th	siebzehnte	27th	siebenundzwanzigste
8th	achte	18th	achtzehnte	28th	achtundzwanzigste
9th	neunte	19th	neunzehnte	29th	neunundzwanzigste
10th	zehnte	20th	zwanzigste	30th	dreißigste
				31st	einunddreißigste

Januar Februar März April Mai Juni Juli August September Oktober November Dezember	Wann hast du Geburtstag? Mein Geburtstag ist am neunundzwanzigsten März Mein Geburtstag ist im März Wann hat deine Schwester/dein Bruder Geburtstag? Er/sie hat am … Geburtstag Am neunzehnten Juni gehen wir in Urlaub fahren wir in Urlaub fliegen wir in Urlaub (auf Urlaub – mainly used in Austria)	den 29. Januar den 13. März

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >Mé féin; An scoil; Ocáidí speisialta • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • Mathematics>Number>Counting and numeration and comparing and ordering	Language Awareness • Examine words which look/sound like the English equivalent (cognates) (e.g. Januar, Februar, März, April, August, September, Okober, November, Dezember) or the Irish equivalent (e.g. Feabhra, Márta) • Examine compound words – neun|zehnten, Geburts|tag • Examine the numbering system for dates across the three languages

• Mathematics>Measures>Time Linkage • Section Three: Numbers 1-20 • Section Four: General Introductions: More Advanced I (birthday) • Section Five: General Introductions: More Advanced II (birthday) • Section Fourteen: Numbers 20+ • Fillers: Months of the year	• Describe nouns, what is a noun? • Examine how all German nouns take a capital letter (e.g. Viertel)

Section Seventeen: Food and shopping

Language:

Was isst du gern? Magst du…(Salat)? Ich mag/esse gern… Ich mag … nicht Ich hasse… Was magst du am liebsten? Ich mag … am liebsten Ich mag … sehr gern	Der Salat Der Fisch Der Kuchen Der Hamburger Der Käse Der Schinken Der Spinat Der Weißkohl Der Blumenkohl Der Rosenkohl Der Pfirsich(e) Der Apfel	Die Pommes Frites Die Pizza Die Butter Die Spaghetti Die Tomate(n) Die Banane(n) Die Traube(n) Die Wurst Die Kartoffel Die Zwiebel(n) Die Erbse(n) Die Himbeere(n) Die Erdberre(n) Die Melone(n) Die Kirsche(n) Die Aprikose(n) Die Pflaume(n) Die Birne(n) Die Apfelsine(n) Die Zitrone	Das Obst Das Hähnchen Das Brot Das Gemüse Das Fleisch

Ins Geschäft Kann ich Ihnen helfen? Das ist teuer Das ist billig	Was kostet das? Das kostet …… Euro Danke schön

Ich möchte Ich will	einen Karton einen Becher eine Packung eine Dose eine Tafel ein Liter ein Kilo ein Pfund ein Glas	Orangensaft Margarine Milch Bohnen Schokolade Milch Kartoffeln Butter Honig	bitte

	ein Stück ein Paket	Kuchen Chips	

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht > Mé féin; Sa bhaile; Bia; Siopadóireacht • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • Mathematics>Number>Counting and numeration and comparing and ordering • Mathematics>Measures>Time; Length; Weight; Capacity; Money • Geography>Human environments>People and other lands (food) • SPHE>Myself>Taking care of my body (Food and nutrition) • History>Change and continuity>Food and farming; Shops; Barter, trade and money Linkage • Section Six: Finding out what things are called • Section Thirteen: Pastimes and Hobbies (I like to cook, shop…) • Fillers: Colour	Language Awareness • Examine words which look/sound like the English equivalent (cognates) (e.g. Salat, Fisch, Hamburger, Pizza, Butter, Spaghetti, Tomate, Pfirsich, Apfel, Banane, Melone, Aprikose, helfen, kosten, Karton, Packung, Liter, Kilo, Pfund, Glas, Paket, Margarine, Schokolade, Milch) or the Irish equivalent (e.g. an burgar, an sailéad, an spaigití, an banana, an t-oinniú, an t-oráiste, cartán, margairín,) • Examine words which look like English words, but which don't relate in meaning (false friends) (e.g. Schinken, ich will, Stück, Chips) • Examine how all nouns in German receive a capital letter. Examine the use of capital letters in English and Irish. • Examine the use of gender with nouns in German (masculine, feminine and neuter) and Irish (masculine and feminine) • Examine compound words – neun|zehnten, Geburts|tag • Examine the numbering system for dates across the three languages • Describe nouns, what is a noun? • Examine how to form the plural of nouns in the three languages • Examine how all German nouns take a capital letter (e.g. Viertel) • Examine how to ask for things politely across the three languages (modal verbs)

Section Eighteen: Fillers

The Alphabet – Das Alphabet

a – ah, apple b - bay c - tsay d - day e – ay, day f - eff g - gay	h – ha, haha i - eeh j - yacht k – kah, cat l - ell m - emm n - enn	o – oh p - pay q – koo, cuckoo r – err, air s - ess t – tae (Ir.), tay u – ooh, poo	v - fow w - vay x – ixx, sticks y – oop-see-lohn, üpsilon z - zett

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language	Language Awareness • Compare the letters and sounds of the alphabet across the three languages. • Irish: a, b c d e f g h i l m n o p r s t u, á, é, í, ó, ú. Letters j, k, q, v, w, x, y and z occur in words borrowed from other languages.

• Music>Listening and responding>Exploring sounds • Geography>Human Environments>People and places in other areas; People and other lands • Science>Energy and forces>Sound Linkage • All raw content – correct pronunciation	• Examine the sounds of the alphabet in the three languages

Days of the week

Montag Dienstag Mittwoch Donnerstag	Freitag Samstag Sonntag

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • Mathematics>Number>Counting and numeration; Comparing and ordering Linkage • Section Eleven: School • Section Thirteen: Pastimes and Hobbies • Section Sixteen: Birthdays and Date	Language Awareness • Examine compound words – Frei|tag, Sonn|tag, Mitt|woch, Donners|tag

Months of the year

Januar Februar März April Mai Juni	Juli August September Oktober November Dezember	Was findet im Dezember statt? Wann fängt die Schule an? Wann gehen/fliegen/fahren wir in Urlaub? Diesen Monat… Letzten Monat… Am 10. dieses Monats…

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >Ócáidí speisialta • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • Mathematics>Number>Counting and numeration; Comparing and ordering	Language Awareness • Examine words which look/sound like the English equivalent (cognates) (e.g. Januar, Februar, März, April, Mai, August, September, Oktober, November, Dezember) or the Irish equivalent (e.g. Feabhra, Márta)

Linkage • Section Eleven: School • Section Sixteen: Birthdays and Date • Fillers: Seasons	

Colours - Die Farbe(n)

Orange Rosa Violett / lila Blau Gelb Grün	Rot Schwarz Braun Grau Weiß

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • Visual Arts>Paint and colour • Science>Energy and forces>Light Linkage • Section Nine: Family II and Describing Yourself I • Section Ten: Describing Yourself II • Section Eleven: School (and uniform) • Section Twelve: Pets • Section Seventeen: Food	Language Awareness • Examine words which look/sound like the English equivalent (cognates) (e.g.orange, rosa, violet, blau, grün, braun, grau) or the Irish equivalent (e.g. Oráiste) • Examine words which look like English words, but which don't relate in meaning (false friends) (e.g. rot)

Shapes

der Kreis der Würfel der Kegel der Zylinder	die Kugel	das Viereck das Dreieck das Rechteck das Oval das Achteck

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • Mathematics>Shape and space>2-D shapes; 3-D shapes • Science>Materials>Properties and characteristics of materials Linkage • Section Seventeen: Food	Language Awareness • Examine words which look/sound like the English equivalent (cognates) (e.g.Oval) or Irish equivalents (e.g. an sféar, an t-ochtagán) • Examine compound words – View|eck, Drei|eck, Recht|eck, Acht|eck •

Seasons - Die Jahreszeit (en)

Der Frühling Der Sommer Der Herbst Der Winter

Integration • Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • Science>Environmental awareness and care>Science and the environment Linkage: • Section Eight: Weather • Fillers: Months of the year	Language Awareness • Examine words which look/sound like the English equivalent (cognates) (e.g. Sommer, Winter) •

Section Nineteen: Classroom Language:

Roll-call	**Praise**	**General**
Hier Nicht hier Ja Nein	Gut Sehr gut Fantastisch Wunderbar Prima toll	Darf ich zur Toilette gehen bitte? Darf ich bitte auf die Toilette gehen? Darf ich bitte aufs Klo gehen? Darf ich meinen Bleistift spitzen? Könnten Sie bitte mal herkommen? Könnten Sie das bitte erklären?

Understanding	**Questions**	**Games**
Ich verstehe das nicht. Ich kapiere das nicht Ich habe keine Ahnung Wiederholen Sie (das), bitte Nochmal/ Noch einmal bitte Wie bitte? Jetzt verstehe ich (das)	Ich habe eine Frage Was machen wir heute? Welche Seite? Wo sind wir? (Auf Seite 2) Was bedeutet 'Freund' auf Englisch Wie sagt man 'see you soon' auf Deutsch? Was ist los? Wer ist das? Bist du sicher? Verstehst du? Wo ist…? Alles klar?	Würfle! Ich bin dran Du bist dran Wer ist dran? Kein Betrügen! Ich habe gewonnen Ich habe verloren Welche Nummer hast du bekommen? Wie viele Karten hast du?

Excuses	**Extras**	**Opinion**
Ich habe meine Hausaufgaben vergessen Ich habe nichts auf. Mein Hund hat die Hausaufgaben gefressen! Was war in Deutsch auf? Haben wir heute Hausaufgaben?	Vielleicht Später Jetzt Danke (schön/sehr) Bitte (schön/sehr) Kein Problem Das macht nichts Nichts zu danken Es tut mir leid Steh auf! Setz dich! Mach die Tür zu! Öffne die Tür! Mach die Tür auf! Halt Sei ruhig!	Das ist einfach/ leicht Das ist schwer Deutsch macht Spaß Spaß muss sein!

Integration	Language Awareness
• Gaeilge>Éisteacht. labhairt, léitheoireacht, scríbneoireacht >téamaí: An Scoil • English> oral language, reading and writing>competence and confidence in using language and receptiveness to language • History>Myself and my family/local studies>My school • SPHE>Myself and the wider world>My school community • Geography>Human Environments>Living in the local community (School) • P.E.>Games Linkage • Section Eleven School • Section Thirteen: Pastimes and hobbies	• Examine words which look/sound like the English equivalent (cognates) (e.g. Hier) • Examine compound words – Haus|aufgaben • Examine how to be polite across the three languages – please, thank you, no problem, I'm sorry • Examine how to negate a sentence across the three languages – not, nicht, ní

Appendix F Sample handouts and worksheets used to present the raw content

The examples provided show materials which were developed specifically around the raw linguistic content, localised supplementary materials which were developed around external resources (poems, songs etc) and materials drafted to match the holiday season (e.g. making pancakes at the beginning of Lent).

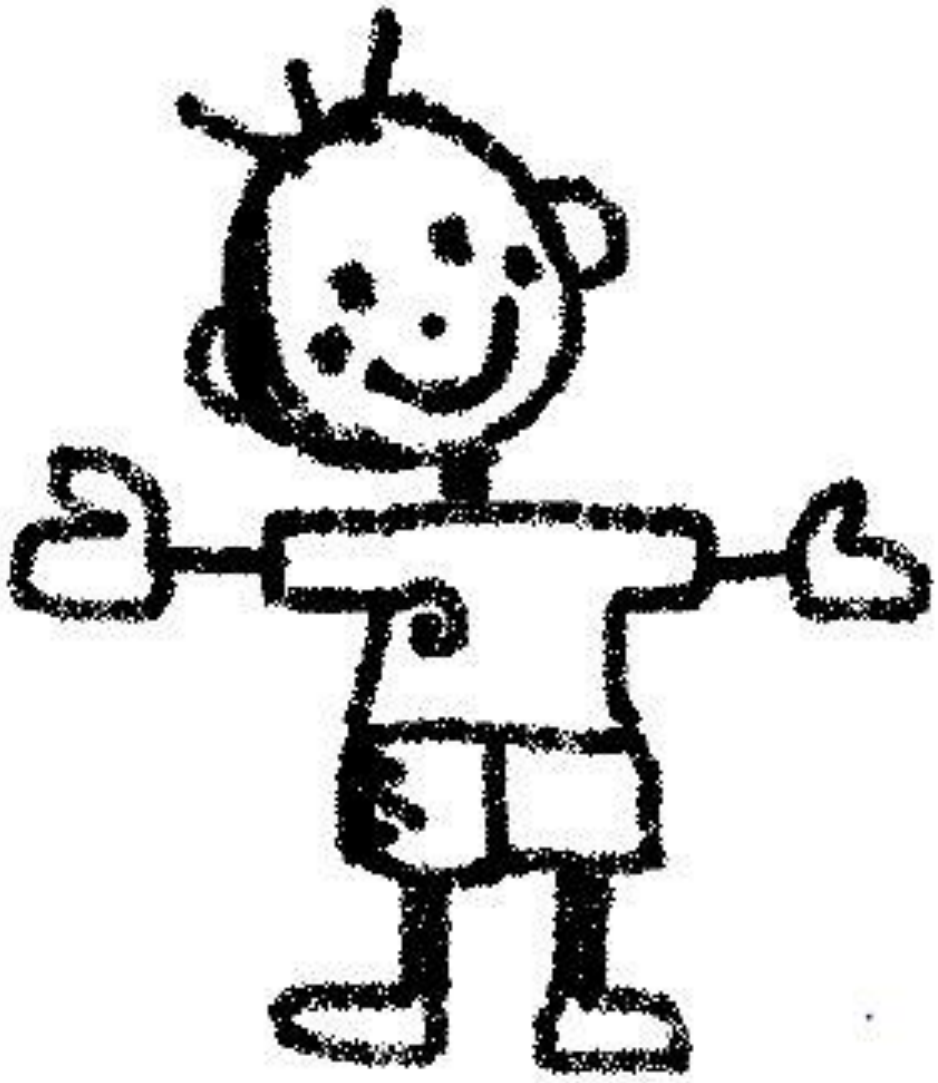
- Those images which were not drawn specifically for the following handouts and worksheets were sourced from Hot Potatoes University of Victoria Language Teaching Clipart Library. Available online at http://hcmc.uvic.ca/clipart/
- References for poetry are provided by the author's byline.
- Sources of other materials (e.g. season's song and music) are provided on the relevant handout or worksheet.

Handout 2 **Body Parts!** **Körperteile!**

Name: _____ Datum: _____

Worksheet 2 **Body Parts!** **Körperteile!**

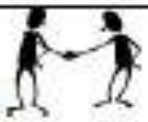
Name: _____ Datum: _____



Der Kopf	Die Haare	Das Auge
Der Daumen	Die Backe	Das Ohr
Der Finger	Die Schulter	Das Bein
Der Bauch	Die Hand	Das Knie
Der Mund	Die Nase	
Der Arm		
Der Fuß		

Handout 4

Greetings II

1. Hello! 

Hallo!
Guten Morgen
Guten Tag
Guten Abend

2. How are you? 

Wie geht's?
Gut, danke! ☺
Sehr gut, danke! ☺
Nicht so gut ☹

3. What is your name? 

Wie heißt du?
Ich heiße _____

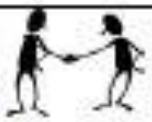
4. How old are you? 10

Wie alt bist du?

Ich bin neun Jahre alt 9
Ich bin zehn Jahre alt 10

5. Where do you come from?

Woher kommst du?
Ich komme aus Irland
Ich komme aus Deutschland
Ich komme aus Frankreich

6. Goodbye! 

Auf Wiedersehen
Tschüs
Bis bald

7. Nationality

Welche Nationalität hast du?

 Ich bin Irin

 Ich bin Ire

Name: _____ Datum: _____

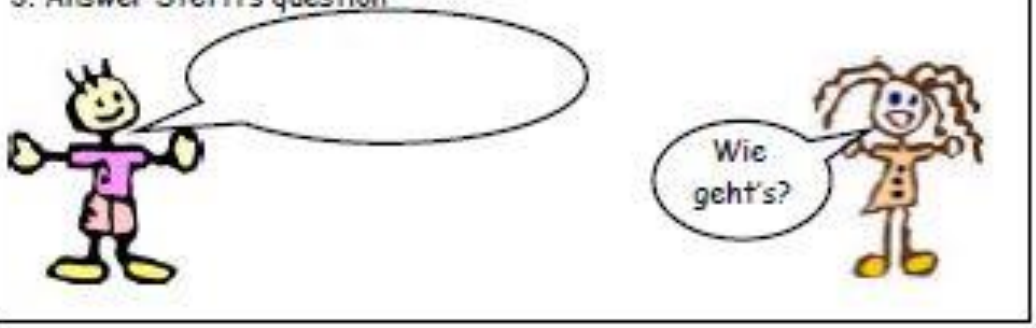
Worksheet 4

Greetings II

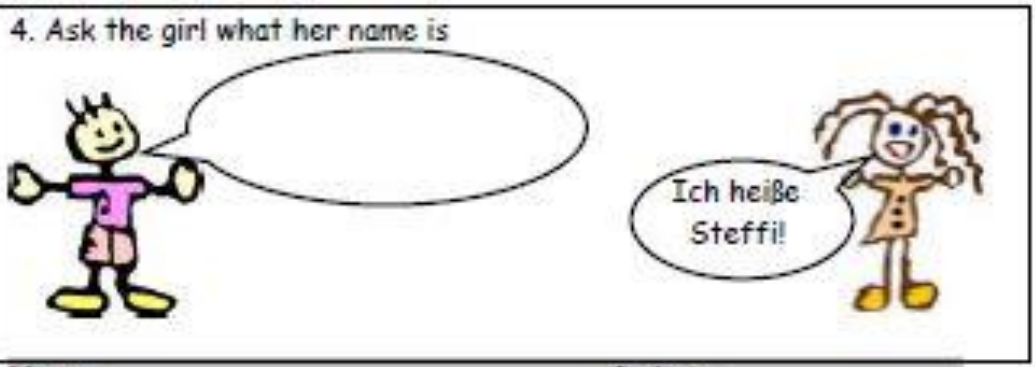
1. Say hello!

2. Ask Steffi how she is

3. Answer Steffi's question



4. Ask the girl what her name is

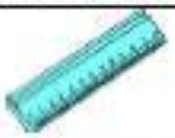


Name: _____ Datum: _____

Handout 6

Was ist das?

Was ist das? Das ist...
Wie heißt das auf Deutsch? Das heißt...

der/ ein	die/ eine	das/ ein
 ein/ der Bleistift	 eine/ die Kreide	 ein/ das Bild
 ein/ der Computer	 eine/ die Lehrerin	 ein/ das Buch
 ein/ der Kuli	 eine/ die Spüle	 ein/ das Fenster
 ein/ der Schrank	 eine/ die Tafel	 ein/ das Glas
 ein/ der Stuhl	 eine/ die Tasche	 ein/ das Heft
 ein/ der Tisch	 eine/ die Tür	 ein/ das Lineal
 ein/ der (Wasser)Hahn	 eine/ die Schere	 ein/ das Regal

NOTE:
*All nouns (person, place or thing) are written with a capital letter
*They can be masculine (der), feminine (die) or neuter (das)

Name: _____ Datum: _____

Worksheet 6

Was ist das?

R R E G A L G I S U Ü E S N					
Y T S X F P Ö F L E F A T Ä					
A R R H Ü R T H I E M E N T					
L E Ü T N E I B N T A P L A					
Ö P T F I T S I E L B P E S					
E O M E A U Ö L A E L A L S					
D N Y N P P T D L H T N Ü R					
W P Ö S U M U Ü G R Ö E P U					
B R E T N O A A R E E M S O					
N T F E H C L Ö E R M H B M					
U A E R R A K U L I U C A U					
F S Ä H E Ü S N R N E F R R					
D C E R M D Ä I S T M E Ü S					
B H Ö B U C H U A Ü F D T T					
I E S U O H E I G O D N A U					
P E A G L Ö Ö R T U H M Ö H					
K N A R H C S Ä G A I D N L					
A C M W G D C T H D O G D N					
N B U Ü T R H R Ö I Ö S E R					
C R H C A L E B K N O I R D					
Ä N G S I S R E A R T Ü N H					
P G L A S T E D M N E I R T					
T C Ö A R S C Ä W Ö M I E I					
F K W O Ü H S E H S S E D L					
Ü N O N E H C S I T Ö H T E					

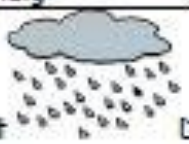
Bleistift	Kreide Bild	Computer	Lehrerin	Buch	Regal
Kuli	Spüle	Fenster	Schrank	Tafel	Glas
Stuhl	Tasche	Heft	Tisch	Tür	Lineal
Schere	Wasserhahn				

Name: _____ Datum: _____

Handout 8

Das Wetter

Wie ist das Wetter heute?

Es ist sonnig  Die Sonne	Es ist wolkig  Die Wolke
Es ist windig  Der Wind	Es ist nebelig  Der Nebel
Es regent  Der Regen	 Es schneit  Der Schnee
Es hagelt  Der Hagel	 Es stürmt Der Sturm (Der Blitz, der Donner: Das Gewitter)

Es ist kalt 	Es ist warm/heiß 	Es ist nass 
 Der Regenschirm	 Der Regenbogen	 Das Eis

Name: _____ Datum: _____

Worksheet 8a

Das Wetter

Exercise 1: Match the correct sentences/words with their pictures

Es ist sonnig			Der Regenschirm
Es regent			Es ist nebelig
Es schneit			Es ist warm
Es hagelt			Der Regenbogen
Es ist wolkig			Es ist windig

Exercise 2: Wie ist das Wetter heute?

a.  ___ ist _____

b.  Es ____ _____

c.  ___ _____

d.  Es ____ _____

e.  ___ _____ _____

Name: _____ Datum: _____

Beim Kaufmann

von James Krüss

Marmelade, Schokolade
kaufen Sie bei mir!
Groben Zucker, feinen Zucker,
alles gibt es hier!

Weiße Knöpfe, schwarze Knöpfe,
Bänder für die Schuh',
Garn und Faden,
gibt's im Laden,
greifen Sie nur zu!

Vogelfutter, Markenbutter,
Mehl und Fett und Grieß,
große Gurken, kleine Gurken,
sauer oder süß!

Tag, Herr Seemann!
Tag, Herr Lehmann!
Womit kann ich dienen?
Meine Eier sind nicht teuer,
die empfehl' ich Ihnen!

Grüne Seife, gelbe Seife,
Ata und Persil!
Kommen Sie und kaufen Sie!
Er kostet gar nicht viel.

Geben Sie mir diese Wurst hier
und ein Viertel Quark!
Gern geschehen, Handumdrehen,
kostet eine Mark!

Dictionary Work: Beim Kaufmann *von James Krüss*

Look up these words in your German←→English dictionary

Der Zucker *the sugar*

Der Knopf

Der Faden

Der Grieß

Der Quark

Die Marmalade

Die Schokolade

Die Butter

Die Gurke

Die Seife

Die Wurst

Das Band

Das Futter

Das Mehl

Das Fett

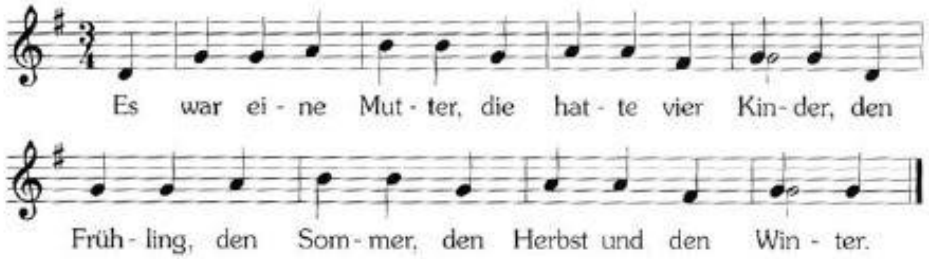
Das Ei





Handout v

Die Jahreszeiten



Es war eine Mutter,
sie hatte vier Kinder,
den Frühling, den Sommer,
den Herbst und den Winter.

Der Frühling bringt Blumen,
der Sommer den Klee,
der Herbst bringt die Trauben,
der Winter den Schnee.

Draw in pictures for the words of the song!

Es war eine Mutter,	sie hatte vier Kinder,
den Frühling, den Sommer, 	den Herbst und den Winter. 
Der Frühling bringt Blumen,	der Sommer den Klee,
der Herbst bringt die Trauben,	der Winter den Schnee.

Source: http://helena.ludwig.name/Kinderwelt/Kinderlieder%20Site/es_war_eine_mutter.htm

Name: _____ Datum: _____

Fastnachtsdienstag/Faschingsdienstag

Was brauchen wir um Pfannkuchen vorzubereiten?!

D	G	K	A	Ö	T	P	F	A	N	N	E
H	U	Ö	L	P	B	L	A	F	Y	I	B
S	C	H	Ö	P	F	L	Ö	F	F	E	L
K	A	T	S	R	M	I	R	M	U	R	A
H	C	L	I	M	H	T	P	E	Z	E	R
H	A	L	S	K	E	R	E	N	I	T	V
C	T	Ö	O	P	W	H	R	L	G	T	W
R	E	T	U	S	N	U	L	R	M	U	A
E	I	E	R	N	E	N	A	N	A	B	N
B	G	O	G	E	N	R	U	Ö	N	I	P
N	E	H	C	U	K	N	N	A	F	P	E
R	T	H	B	R	E	K	C	U	Z	B	Ö

Mehl	Milch	Pfannkuchen	Rezept
Butter	Zucker	Öl	Teig
Eier	Pfanne	Schöpflöffel	Bananen

Das Rezept: ZUTATEN

250g Mehl	4 Eier
50g Butter (oder Sonnenblumenöl)	½ l Milch

Die Eier gesondert aufschlagen, damit keine Schalen in den Teig kommen. Die Eier, das Mehl und die Milch in eine Schüssel geben. Umrühren. Wenn der Teig zu fest ist, noch etwas Milch dazugeben.

Eine Pfanne heiss machen, dann ein wenig Butter oder Öl hinein. Mit einem Schöpflöffel so viel Teig hineingießen, dass gerade der Boden bedeckt ist. 1½ Minunten braten, dann wenden und noch mal ½ Minuten.
Mit Zucker, Nutella, Bananen, Marmalade oder irgend etwas anderem servieren oder füllen.

Name: _____ Datum: _____

Appendix G Structured feedback form for data gathering on raw content

(Note: Numbering added during analysis for easier cross-referencing)

1. Date	
2. Class	
3. Raw content topic	
Cross reference (4) h/o and (5) w/s	

Content

1. Missing language? <introduced through the course of the session>
2. Superfluous language?
3. Relevance of language to children Not relevant □ Appropriate □ Very relevant □ 3a. Comment:
4. Any difficulties encountered? <vocabulary, pronunciation etc.> Yes □ No □ 4a. Detail:

Resources/Presentation

5. Clarity of relevant h/o and w/s Not clear □ Appropriate □ Very clear □ 5a. Comment:
6. Age-appropriateness of h/o and w/s Not appropriate □ Appropriate □ Very appropriate 6a. Comment:
7. Any changes required? Yes □ No □ 7a. Detail:

8. Researcher's ancillary notes

Appendix H Error Analysis: Closed-ended checklist of learner error domains and categories

Error domain		Error category		Error type		Description/detail
<W>	Written form	<FM>	Form	<CA>	Incorrect case (upper/lower)	All nouns upper case
				<DIA>	Diacritic (missing or erroneous)	German umlaut missing or in incorrect location
				<PS>	Phonetic spelling	Using knowledge of English or Irish sounds to spell German words as they sound (e.g. drei/ dry)
				<SE>	Spelling Error (Misc.)	Other spelling error
		<PN>	Punctuation	<MP>	Missing punctuation	
				<EP>	Extra punctuation	
<L>	Language/Lexis	<CS>	Code switching	<ENG>	Use of English word	Use of English word in place of German one, e.g. 'sister' for 'Schwester'
				<IR>	Use of Irish word	Use of Irish word in place of German one, e.g. 'naoi' for 'neun/nine'
		<CE>	Choice	<IC>	Inappropriate lexical choice	e.g. use of 'ich/I' where 'er/he' was intended
<M>	Morphology	<IN>	Inflection	<PL>	Incorrect plural form/Missing plural marker	e.g. zwei Schwester*/*two sister** should include the plural marker 'n' at the end of 'Schwester'
				<II>	Incorrect inflection	e.g. verb conjugation, genitive case marker on nouns
				<MI>	Missing inflection	
<G>	Grammar	<AT>	Agreement	<ArtN>	Art N	Gender/case agreement
				<AdjN>	Adj N	Number/case agreement
				<SubjV>	Subj V	Person agreement
				<AdvN>	Adverb N	Gender case agreement
				<PrepArtN>	Preposition Article Noun agreement	Gender/case agreement
<S>	Syntax			<WO>	Word order	Incorrect word order
				<MW>	Missing word	
				<EW>	Extra word	e.g. Es *ist* regnet* for Es regnet/it is raining

Error domain		**Error category**	**Error type**		**Description/detail**
<U>	Understanding		<Und>	Lack of understanding	Lack of understanding during oral conversation/giving directions
<P>	Phonology		<Pron>	Difficulties with pronunciation	During oral test/conversation/practice

Appendix I Open-ended comments recorded on the Structured Feedback Form

1. Missing Language

- Frau Keogh, Kindergarten, Autobahn, Burger – Hamburger (Hamburg), Frankfurter – Frankfurt, Kaput, Rucksack – back-pack, Pretzel – from German Brezel, Darf icg zur Toilette gehen bitte?
- Zungenbrecher - tongue twister, Hausaufgaben
- die Fahne, Deutshland, Österreich, die Schweiz, Lichtenstein, Ich spreche Irisch, Englisch, Deutsch, Französisch, Spanisch
- Katze, klug, klein, kratzen, krokodile
- Quiz
- gelb, Rübe
- Polizei, ein Glas Bier, alt, Hex
- Bingo
- Berlin, München, Zürich, Stuttgart,
- Der Esel, Das Lamm, wecken, der Weihnachtsmann, das Schwein, mitnehmen
- der Engel, die Gloecke, die Kerze, die Krippe, Weihnachten, Weihnachtskarten, der Stern,
- Was ist das? Wie heißt das auf Deutsch? Was bedeutet …
- Wie komme ich zum/zur, der Bahnhof, der Markt, das Schwimmbad, die Schule, die Kirche, das Hotel, das Rathaus, die Geschaefte, das , gerade aus, um die Ecke, Strasse, links, rechts, erste, zweite
- Welche Telefonnummer hast du? Meine Telefonnummer ist …, Wer hat …, zeig mir , weather
- die Kamera, der Schluessel, die Uhr, der Stadtplan, das Buch, der Kuli
- Er kommt aus Schottland, Indien, Bonn, Hamburg, München
- Plus, minus
- Was bin ich? Schlange, Regenschirm, Auto, Haare, Regenbogen, Wolke, Wasser, Bauch
- der Löwe, Tiger, Giraffe, der Hund
- Mehl, Milch, Ei,, Öl
- Das Pferd, der Panda, die Ziege, Das Zebra, Der Gorilla, Die Giraffe, Der Tiger
- Der Henker, du bist tot,
- Frohe OsternEi, Osterhase
- Die Nase ist …, Der Mann ist ….
- Am X fahren wir im Urlaub.
- Das Auto ist…, Der Schneemann ist…, Die Wolke ist…, Die Sonne ist…, Das Wasser ist…,, Die Blume ist …, Die Schuhe sind...
- England, Afrika, Frankreich, Spanien
- und, oder, aber, auch
- Zeig mir
- der Mond, die Sonne, die Mitte, der Donner
- Mantle, Regenschirm, sei klein, sei gross
- Kunst
- Sanniklaus, Stille Nacht
- Teig, Pfanne, Pfannkuchen, Eier aufschlagen
- der Frohe Ostern, Fruehling Osterhase, das Osterei, das Osterfest, die Osterlocke (daffodil), der Ostersonntag, das Ostern
- Noch etwas? Das ist billig/teuer
- Die Milch, Die Muschel, der Krebs, Teebeutel, Karotte, Pizza, Kase, Kekse, Suppe, Hamburger, Seife, Zahnpasta, Haarwaschmittel
- Ich hasse, ich mag, essen, trinken
- auf dem Land wohnen, in der Stadt leben, sauber, schmutzigreich, arm, eines Tages, besuchen, die Reise, das Flugzeug, der Krankenwagen, der Bus, das Motorrad, das Fahrrad
- der Himmel, die Plaetze, das Eis, der Sperling, Dach
- Zimmer, Buero, Ich verstehe nicht
- der Kuerbis (pumpkin), Fasching, Fastnacht
- Sauerkraut, Bier, Berg, ja, nein, alle fertig, Herr

- Nord, Sud, Ost, West
- Ich bekomme 5 Taschengeld pro Woche

3a. Relevance of content to children: Comment

No comments recorded

4a. Any difficulties encountered: Comment

- Many words sound like English equivalents
- Understanding - more explanation required and assistance with vocabulary and pronunciation
- Poem good for reinfrcing raw content
- Many words - a little overwhelming
- A lot of vocab to cover off. Pages are quite dense
- Mathematical concept hard for them to grasp, to reverse numbers
- Instructions for preparing the pancakes were a little challenging. Not many known words among them.
- Prob too much vocab on one handout, spread across a second handout.
- Back to numbers issue…need to crack 21+ before dates will be clear, although children are just memorising their birthday and don't worry about the rest.
- Too many words for one sitting. Need to gradually introduce them
- A lot on each page - densely packed

5a. Clarity of relevant h/o and w/s: Comment

- Graphical representation to aid understanding. Tied in with Irish parts of the body exploration
- More detailon compounding and making up numbers 20+
- Needs to be filtered by students for information

6a. Age-appropriateness of h/o and w/s: Comment

- Needs adapting
- Perhaps a little old (getting a job and Gastarbeiter)
- Students filter relevant content

7a. Any changes required: Detail

- Maybe break down poems into smaller parts to allow for understanding. Also a 'Wortschatz' to go along with it.
- Less detail or dictionary provided alongside. Or simplified version of text provided.
- Less on each page - overwhelms children to have too much on one page
- More images for the recipe method. Wortschatz to go with it.
- Adapt across more than one handout
- Less content per page
- Less per page, spread across more pages

8. Researcher's ancillary notes

- Children want to learn about games, music, songs, artists, poems, flags, football, flashcards, siopa, body parts, how to write in German
- Extra content from non-raw materials
- Starting with numbers 1-10 also
- Much revision required of already covered material
- Halloween poem, good for variety in resource material
- Poetry to revise content learner. Lots of revision required

- All content is revised and built upon as we go. Also work on role play - about myself and body parts - Pumpernickel sagt
- Only learning numbers 11-20 this class
- Importance of tying in German teaching with what's going on in the school and holiday time - Xmas
- Children wanted to know names of places in Germany to answer 'Woher kommst Du?'
- Children's pronunciation of poetry is good, looked at Muede poem again
- Children eager to find out vocabulary items in German for Xmas
- Introduced telephone numbers
- Classroom directions for showing large number flashcards.
- Homework all spelt phonetically – e.g. drei/dry
- Children finding it hard to put numbers together in reverse. Not good to draw too many parallels with other languages until they comprehend and acquire the German equivalent.
- Hangman, alphabet to music worked well
- Spread handout over multiple classes and used as a point of reference for homework etc
- Good to compile all relevant information into one place for revision purposes.
- Singular and plural - ist and sind.
- Use of Usbourne sheets to revise content with alt resource
- Combining sentences on weather. How to remember pronunciation of words - Tisch, hink of tissue on table.
- Still working on classroom words, introduced too many in one sitting initialy
- Trying to make new sentences with all known vocab - e.g. Welche Farbe ist die Tür
- Asking what they would like to learn about this term. Similar to last year, with added songs, TV, school system, olympics, football, dodgeball, rugby, German sports. Prep for project and looking at postcard with Hello in lots of languages
- Looked at parts of speech - noun, verb and adjective
- Break down days of the week into other words
- Seasons
- Children like describing girlfriends, and calling people fat etc
- Clothes we need for certain weather
- Guess who in German. Describe a person, we have to guess wh
- Classroom language cards
- Copy check - talked about writing full sentences, translating sentences into German and English, capital letters
- ing)Pancake ingredients and toppings, pictures and packages from food (to look at contents listn
- Pancake ingredients and toppings, pictures and packages from food (to look at contents listing), umlauts
- Role play, ins Geschaft
- Behaviour difficulties, children taken out to help set up for ST. Patrick's Day parade.
- Role-play revision - needs constant work
- Mistakes with capitals, umlauts, articles, spelling long words, ie/ei mix up
- Mistakes with der/die/daas, spelling, umlaut, capital letters
- A lot of writing today
- Good for learning off fixed phrases
- Number revision
- Halb X presenting some difficulties
- Not too much knowledge of Geography
- Good retention generally, but revision required regularly

Appendix J Oral Test Preparation and Questions

Oral Test Information

These are the **exact** questions you will be asked.

Use **Role-Play Summary 2** to help you revise.

1. Wie sagt man '*hello*' auf Deutsch?	How do you say '*hello*' in German?
2. Wie geht's?	How are you?
3. Wie heißt du?	What is your name?
4. Wie alt bist du?	How old are you?
5. Woher kommst du?	Where do you come from?
6. Wo wohnst du?	Where do you live?
7. Hast du Geschwister?	Do you have any brothers or sisters?
8. Wann hast du Geburtstag?	When is your birthday?
9. Was ist deine Telefonnummer?	What is your telephone number?
10. Hast du Haustiere?	Do you have any pets?
11. Wie ist das Wetter heute?	What is the weather like today?
12. Wie sagt man '*goodbye*' auf Deutsch?	How do you say '*goodbye*' in German?

Appendix K Student errors recorded during school-based research

Error domain		**Error category**		Error type		Count (n)	Percentage (%)	Subtotal
<W>	Written form	<FM>	Form	<CA>	Incorrect case (upper/lower)	1391	20	
				<DIA>	Diacritic (missing or erroneous)	324	5	
				<PS>	Phonetic spelling	422	6	
				<SE>	Spelling Error (Misc.)	448	6	
		<PN>	Punctuation	<MP>	Missing punctuation	45	1	
				<EP>	Extra punctuation	2	0	**2632 (37%)**
<L>	Language/Lexis	<CS>	Code switching	<ENG>	Use of English word	473	7	
				<IR>	Use of Irish word	16	0	
		<CE>	Choice	<IC>	Inappropriate lexical choice	1	0	490 (7%)
<M>	Morphology	<IN>	Inflection	<PL>	Incorrect plural form/Missing plural marker	120	2	
				<II>	Incorrect inflection	58	1	
				<MI>	Missing inflection	17	0	195 (3%)
<G>	Grammar	<AT>	Agreement	<ArtN>	Article Noun agreement	752	11	
				<AdjN>	Adjective Noun agreement	157	2	
				<SubjV>	Subject Verb agreement	554	8	
				<AdvN>	Adverb Noun agreement	103	1	
				<PrepArtN>	Preposition-Article Noun agreement	106	1	1672 (24%)
<S>	Syntax			<WO>	Word order	299	4	
				<MW>	Missing word	237	3	
				<EW>	Extra word	174	2	710 (10%)
<U>	Understanding			<Und>	Lack of understanding	519	7	519 (7%)
<P>	Phonology			<Pron>	Difficulties with pronunciation	857	12	857 (12%)
TOTAL						7075	99	7075 (100%)

(Note: Percentages rounded to the nearest integer)

Appendix L Extract from raw linguistic content v1.0 in .txt format after transfer from MS Word

Sections 1 - 4 are included in this extract

// NOTES: // [] insert choice. e.g. Ich heiße [Klaus].
// All single nouns end in a full-stop too.
// Topic choices start with "Section 'number in words'".
// Topics under main headings have a * before them.

Section One: General Introductions.

* 1. Greetings.

Hallo! Guten Morgen. Guten Tag. Guten Abend. Gute Nacht. Schlaf gut. Auf Wiedersehen.
Tschüs. Bis bald.
Bis morgen.
Bis Donnerstag.

* 2. How are you?

Wie geht's?
Gut, danke! Sehr gut, danke! Nicht so gut.
Schlecht.
Mir geht's gut.
Mir geht's sehr gut.

* 3. What is your name?
Wie heißt du? Ich heiße [John].

Section Two: Parts of the body.

Der Bauch.
Der Daumen.
Der Finger.

Der Fuß.

Der Hals.

Der Kopf.

Der Mund.

Der Rücken.

Die Backe.

Die Haare (plural).

Die Hand.

Die Nase.

Die Schulter.

Die Zehe.

Die Zunge.

Das Auge.

Das Bein.

Das Kinn.

Das Knie.

Das Ohr.

Section Three: Numbers.

Null.

Eins.

Zwei.

Drei.

Vier.

Fünf.

Sechs.

Sieben.

Acht.

Neun.

Zehn.

Elf.

Zwölf.

Dreizehn.

Vierzehn.

Fünfzehn.

Sechzehn.

Siebzehn.

Achtzehn.

Neunzehn.

Zwanzig.

Die Zahlen.

Was ist deine Telefonnummer/ Rufnummer?

Welche Telefonnummer/ Rufnummer hast du/haben Sie?

Meine Telefonnummer/Rufnummer ist 01 6749375.

Ich bekomme €5 Taschengeld pro Woche.

Section Four: General Introductions - more advanced 1.

* 1. How old are you?

Wie alt bist du? Ich bin neun/ zehn Jahre alt.

* 2. Where do you come from?

Woher kommst du? Ich komme aus [Irland/ Deutschland/ Frankreich/ den USA/ Österreich/ der Schweiz].

* 3. (Welcher Nationalität bist du?)

Welche Nationalität hast du?

Welche Staatsangehörigkeit hast du?

Ich bin Irin (fem.).

Ich bin Ire (masc.).

Die Iren.

Appendix M TreeTagger tagset

English translation and examples sourced from Durrell, Bennett, Scheible and Whitt (2012)

POS	Description	Example
ADJA	attributive adjective (including participles used adjectivally)	das **große** Haus die **versunkene** Glocke
ADJD	predicate adjective; adjective used	der Vogel ist **blau** er fährt **schnell**
ADV	adverb (never used as attributive adjective)	sie kommt **bald**
APPR	preposition left hand part of double preposition	**auf** dem Tisch **an** der Straße entlang
APPRART	preposition with fused article	**am** Tag
APPO	postposition	meiner Meinung **nach**
APZR	right hand part of double preposition	an der Straße **entlang**
ART	article (definite or indefinite)	**die** Tante; **eine** Tante
CARD	cardinal number (words or figures); also declined	**zwei; 526; dreier**
FM	foreign words (actual part of speech in original language may be appended, e.g. FM- ADV/ FM-NN)	**semper fidem**
ITJ	interjection	**Ach**!
KON	co-ordinating conjunction	**oder** ich bezahle nicht
KOKOM	comparative conjunction or particle	er arbeitet **als** Straßenfeger
KOUI	preposition used to introduce infinitive	**um** den König zu töten
KOUS	subordinating conjunction	**weil** er sie gesehen hat
NE	names and other proper nouns	**Moskau**
NN	noun (but not adjectives used as nouns)	der **Abend**
PAV	pronominal adverb	sie spielt **damit**
PDAT	demonstrative determiner	**dieser** Mann war schlecht
PDS	demonstrative pronoun	**dieser** war schlecht
PIAT	indefinite determiner without determiner	**kein *Mensch*** **irgendein** Glas
PIDAT	indefinite determiner with determiner	Ein **wenig** Wasser Die **beiden** Brüder
PIS	indefinite pronoun	sie hat **viele** gesehen
PPER	personal pronoun	**sie** liebt **mich**
PRF	reflexive pronoun	ich wasche **mich** sie wäscht **sich**
PPOSS	possessive pronoun	das ist **meins**
PPOSAT	possessive determiner	**mein** Buch das ist der **meine/meinige**
PRELAT	relative pronoun depending on a noun	der Mann, **dessen** Lied ich singe [...], **welchen** Begriff ich
PRELS	relative pronoun (i.e. forms of *der or welcher*)	der Herr, **der** gerade kommt
PTKA	particle with adjective or adverb	**am** besten, **zu** schnell, **aufs** herzlichste
PTKANT	answer particle	**ja, nein**
PTKNEG	negative particle	**nicht**
PTKVZ	separable prefix	sie kommt **an**
PTKZU	infinitive particle	**zu**

PWS	interrogative pronoun	**wer** kommt?
PWAT	interrogative determiner	**welche** Farbe?
PWAV	interrogative adverb	**wann** kommst du?
TRUNC	truncated form of compound	**Vor-** und Nachteile
VAFIN	finite auxiliary verb	sie **ist** gekommen
VAIMP	imperative of auxiliary	**sei** still!
VAINF	infinitive of auxiliary	er wird es gesehen **haben**
VAPP	past participle of auxiliary	sie ist es **gewesen**
VMFIN	finite modal verb	sie **will** kommen
VMINF	infinitive of modal	er hat es sehen **müssen**
VMPP	past participle of auxiliary	sie hat es **gekonnt**
VVFIN	finite full verb	sie **ist** gekommen
VVIMP	imperative of full verb	**bleibt** da!
VVINF	infinitive of full verb	er wird es **sehen**
VVIZU	infinitive with incorporated *zu*	sie versprach **aufzuhören**
VVPP	past participle of full verb	sie ist **gekommen**
XY	Non-word, special character/symbol	D2XW3
$,	Comma	,
$.	Sentence-final punctuation	. ? ! ; :
$(	Other sentential characters or brackets	- [ ] ( )

Appendix N Automatically creating comparative data sources: Perl script listing and associated code

File Ref.	File name	Description
1	Convert_corpus_to_pure_text2.pl	written to extract POS information from a POS-tagged file and replace each token within its full sentence on its own line
2	Convert_words_to_sent_per_line.pl	Written to replace each token within its full sentence on its own line
3	Tiger_preprocessing.pl	Written to extract the token and POS information from the TIGER treebank
4	get_consecutive_sample3.pl	Written to create consecutive random sample files from the edited TIGER corpus (token+POS from 3 above)
5	Get_random_sample.pl	Written to create random sampe files (random selections of words together) from the edited TIGER corpus
6	Remove_hash_text_in_textbook.pl	Remove all hashtags and associated comments from the Passwort Deutsch textbook

Ref: 1

Filename: convert_corpus_to_pure_text2.pl

```
#!/usr/bin/perl

# difference to convert_corpus_to_pure_text.pl in that it write the token out one per line

#print "Hello, World...\n";

#Extract pure text from file and write it to a new file with one word per line
#Try to adjust afterwards to having one sentence per line - separate by punctuation.

open(F, $ARGV[0]) || die("Enter a filename to open");

while ($line = <F>) {
    chomp($line);
    push(@words , (split(/[\t]/, $line)));  #split each line of the corpus at the tab and add each
                                            # of the three elements to the array @words
}   #end while
close(F);

#print ("\n @words \n");

# ARRAY @words holds the corpus: $i is the word, $i+1 is the POS



open(NEW, ">$ARGV[1]") || die("Enter a filename for the new file");

for (my $i=0; $i <= $#words ; $i+=3) {
    #print("$words[$i]");
    push(@just_words, $words[$i]);
}

for (my $k=0; $k<=$#just_words; $k++) {
    print (NEW "$just_words[$k]\n");
}
close(NEW);

#print ("\n @just_words \n");
# Array @just_words holds the pure text with no POS or lemma information

```

Ref: 2

Filename: Convert_words_to_sent_per_line.pl

```
#!/usr/bin/perl

# convert_words_to_sent_per_line.pl

# Convert files with one word per line to one sentence per line

# open the file with one word per line
open(F, $ARGV[0]) || die("Enter a filename to open");

while ($line = <F>) {
    chomp($line);
    push(@words, $line);
}   #end while
close(F);

#print ("\n @words \n");

# ARRAY @words holds the corpus: $i is the word


# open file to write out to
open(NEW, ">$ARGV[1]") || die("Enter a filename for the new file");


for (my $k=0; $k<=$#words; $k++) {
    if (($words[$k-1] eq '.') || ($words[$k-1] eq '?') || ($words[$k-1] eq '!')) {

                                        #if previous element is end of file,
                                        # print out new line before next element
        print(NEW "\n");
    }
    print (NEW "$words[$k] ");
}

close(NEW);


```

Ref: 3

Filename: tiger_preprocessing.pl

```
#!/usr/bin/perl

#TiGer Pre-Processing

my($line);

$word;
$three_tabs;
$pos;
$rest;

# Read in the file
open(TIGER, $ARGV[0]) || die("Enter a filename to open");

while ($line = <TIGER>) {
        chomp($line);
        #print("\n\nline read in is: $line");

    if (!($line=~/^(\#|\%)/)) {         #if the line starts with a '#' or '%'-
        chomp($line);                           #/^\#|\%/
        #print("\nline without # or % is: $line");

        $line=~/(\S+)(\t+)(\S+)((\w|\t)*)/g;
        #print("\nword is:$1 pos is:$3");

        push(@tiger_words, $1, $3); #add word and pos to new array
        }   #end if - do nothing

}   #end while
close(TIGER);

#print ("\n @tiger_words \n");

# ARRAY @tiger_words holds the corpus: $i is the word, $i+1 is the POS


#Write out word and pos to the array
open(NEW_TIGER, ">$ARGV[1]") || die("Enter a filename for the new file");


for (my $i=0; $i <= $#tiger_words ; $i++) {
    #print("$tiger_words[$i]");
    push(@tiger_trimmed, $tiger_words[$i]);
}

#print out trimmed down file with a word and pos per line separated by a tab
for (my $j=0; $j<=$#tiger_trimmed; $j+=2) {
    print (NEW_TIGER "$tiger_trimmed[$j]\t$tiger_trimmed[$j+1]\n");
}

close(NEW_TIGER);

```

Ref: 4

Filename: get_consecutive_sample3.pl

```
#!/usr/bin/perl

# get_consecutive_sample2.pl
# To run: perl get_consecutive_sample2.pl tiger_trimmed.txt 'file_to_write_out_to.txt'

# for consecutive text

# Generates random sampling subtexts from the Trimmed Tiger corpus
# Random samples are 4013 lines long (same as curriculum corpus for comparison)

#

####################
# Get random number #
####################

$random_number = int (rand 708304) +1;  # get random number between 1 and 712319
                                        # Total no. lines in TiGer trimmed file is 712319
                                        # elements in an array will be saved with two indexes/line
                                        # 1424638-8030 (4013*2=8026: the number of array places the sample must be)
                                        # subtracting from the total number of array elements in tiger corpus
                                        # ensures we don't run off the end of the file

                                        #712319-4015=708304
#must start at an even index
# therefore random number*2 will ensure it's even.

$random_number=$random_number*2;

#$random_number = int (rand 24) +1;
#print ("\nrandom number is:$random_number");

##########################################
# Read in trimmed Tiger file to array @words #
# and put each line into two array indexes   #
##########################################

open(TIGER_TRIMMED, $ARGV[0]) || die("Enter a filename to open");


while (my $line = <TIGER_TRIMMED>) {
    chomp($line);
    push(@words , (split(/[\t]/, $line)));  #split each line of the corpus at the tab and add each
                                            # of the three elements to the array @words
}   #end while
close(TIGER_TRIMMED);

#############################################################################################
# Look in array @words at random point                                                      #
#  take in the file from there for count of 4013                                            #
# I am aware that it could also be random number+1 to account for the way arrays are indexed #
#############################################################################################

$sample_index=$random_number+8025;  # int here will be 4013*2=8026

#must have even starting index and odd ending index to have them matching

for (my $i=$random_number; $i<=$sample_index; $i++) {   # start reading at the random index
    push(@con_sample, $words[$i]);

    }   #end for

#################################
# write out sample to a new file #
#################################

open(NEW_TIGER, ">$ARGV[1]") || die("Enter a filename for the new file");

for (my $k=0; $k<=$#con_sample; $k+=2) {
    print (NEW_TIGER "$con_sample[$k]\t$con_sample[$k+1]\n");
}

close(NEW_TIGER);

```

Ref: 5

Filename: get_random_sample.pl

```
#!/usr/bin/perl

# get_random_sample.pl
# To run: perl get_random_sample.pl tiger_trimmed.txt 'file_to_write_out_to.txt'
# For me: perl get_random_sample.pl tiger_trimmed.txt ran.sample1-10.txt

# for random words totallling 4013 tokens

# Generates random sampling subtexts from the Trimmed Tiger corpus
# Random samples are 4013 lines long (same as curriculum corpus for comparison)
# samples are random words totalling 4013 lines, NOT CONSECUTIVE TEXT

###################################
# Get set of random numbers       #
# Write out to array @random_nums #
###################################

# get 4013 random numbers

for (my $j=1; $j<=4013; $j++) {     #sld be 4013
    $random_number = int (rand 708304) +1;  # Total no. lines in TiGer trimmed file is 712319
    #$random_number = int (rand 12) +1;

    # elements in an array will be saved with two indexes/line
    # 1424638-8030 (4013*2=8026: the number of array places the sample must be)
    # subtracting from the total number of array elements in tiger corpus
    # ensures we don't run off the end of the file
    # 712319-4015=708304

    $random_number=$random_number*2;        #must start at an even index
                                            # therefore random number*2 will ensure it's even.
    #print("\nRandom number is:$random_number");
    push(@random_nums, $random_number); # push numbers onto a new array: @random_nums
}   # end for

# Can't repeat any of the random numbers

for (my $x=0; $x<=$#random_nums; $x++) {    #loop through new array @random_nums
    for (my $z=$x+1; $z<=$#random_nums; $z++) {

        if ($random_nums[$x]==$random_nums[$z]) {
            $random_number = int (rand 12) +1;
            $random_nums[$x]=$random_number*2;
        }   #end if
    }   #end for
}   #end for

#print("\nnew random number array is @random_nums");
###############################################
# Read in trimmed Tiger file to array @words  #
# and put each line into two array indexes    #
###############################################

open(TIGER_TRIMMED, $ARGV[0]) || die("Enter a filename to open");

while (my $line = <TIGER_TRIMMED>) {
    chomp($line);
    push(@words , (split(/[\t]/, $line)));  #split each line of the corpus at the tab and add
                                            # each of the three elements to the array @words
}   #end while
close(TIGER_TRIMMED);

#############################################################################################
# Look in array @words at random point                                                      #
#  take in the file from there for count of 4013                                            #
# I am aware that it could also be random number+1 to account for the way arrays are indexed #
#############################################################################################

#$sample_index=$random_number+8025; # int here will be 4013*2=8026

#must have even starting index and odd ending index to have them matching

for (my $f=0; $f<=$#random_nums; $f++) {        #for each random number

    $index=$random_nums[$f];
    push(@random_sample, $words[$index], $words[$index+1]);

}   #end for
```

Ref: 6

Filename: remove_hash_text_from_textbook.pl

```
#!/usr/bin/perl


# File to remove all publisher/author comments from textfile
# and write the result out to a new file
# All comments are surrounded by set of #Comment# and only
# run on one line

# Open file
open(FILE, $ARGV[0]) || die("Enter a filename to open");

while ($line = <FILE>) {
    chomp($line);
    $line =~ s/\#[^\#]*\#//g;   #remove "#comments#"
    push(@words ,$line);        #push line onto an array
}   #end while
close(FILE);

#Write it out to a new file
open(NEW, ">$ARGV[1]") || die("Enter a filename for the new file");

for (my $k=0; $k<=$#words; $k++) {
    print (NEW "$words[$k]\n");
}
close(NEW);

```

Appendix O Automatically processing the datasets and extracting content for SLA ratio calculation: Perl script listing and associated code

File Ref.	File name	Description
7	Find_store_adj_noun_verb.pl	Finds all of the adjectives, nouns and verbs in a file and outputs each of the word category listings to their own file (e.g. adjectives tokens, noun tokens etc.)
8	Count_verb_types_from_tokens.pl	Counts the number of types of a particular word class (e.g. verbs) from a listing of relevant tokens.
9	Count_compare_to_freq_lists_output_tokens.pl	Compares the content of a dataset to a defined frequency list input file. Outputs the sophisticated and basic tokens (for use with file 8 to calculate types from tokens)

Ref: 7

Filename: Find_store_adj_noun_verb.pl

```
#!/usr/bin/perl


# find_store_adj_noun_verb.pl
# To run: perl -w find_store_adj_noun_verb.pl 'sample file with data[word+POS]' 'new file for adjs'
#         'new file for nouns' 'new file for all verbs' 'new file for closed-class words'


# Read in the file
# (1) Find the adjs (ADJA, ADJD)
# (2) Find the nouns (NN, NE)
# (3) Find the modal and auxiliary verbs (VAFIN, VAIMP, VAINF, VAPP, VMFIN, VMINF, VMPP)
# (4) Find the other verbs (VVFIN, VVIMP, VVINF, VVIZU, VVPP)
# Count number or each type found
# write out the elements to a 5 files (to evaluate using wordsmith tools)
# File 1: adjs; File 2: nouns; File 3: all verbs; File 4: adjs & nouns & other verbs (closed_class)


# Read in file

open(SAMPLE, $ARGV[0]) || die("Enter a filename to open");

while (my $line = <SAMPLE>) {
    chomp($line);
    push(@words , (split(/[\t]/, $line)));  #split each line of the corpus at the tab and add each
                                            # of the three elements to the array @words
}   #end while
close(SAMPLE);

#print("Read in data file successfully\n");
#Find the conjunctions (file is format WORD TAB POS, so CONJS can be found at all odd indexes)

for (my $i=0; $i<=$#words; $i+=2) {

    #find adjs
    if (($words[$i+1] eq 'ADJA') || ($words[$i+1] eq 'ADJD')) {     # if the POS is adj
        push(@adjs, $words[$i]);        # write the word out to the array @adjs
    }

    # find nouns
    if (($words[$i+1] eq 'NN') || ($words[$i+1] eq 'NE')) {     # if the POS is noun
        push(@nouns, $words[$i]);       # write the word out to the array @nouns
    }

    # Find aux and modals
    if (($words[$i+1] eq 'VAFIN') || ($words[$i+1] eq 'VAIMP') || ($words[$i+1] eq 'VAINF') || ($words[$i+1] eq 'VAPP') ||
        push(@verbs_a_m, $words[$i]);       # write the word out to the array @verbs_a_m
    }

    # Find other verbs
    if (($words[$i+1] eq 'VVFIN') || ($words[$i+1] eq 'VVIMP') || ($words[$i+1] eq 'VVINF') || ($words[$i+1] eq 'VVIZU') |
        # if the POS is noun
        push(@verbs, $words[$i]);       # write the word out to the array @verbs
    }
}   #end for

# Write out to 4 new files - one token per line
# File 1: adjs; File 2: nouns; File 3: all verbs; File 4: adjs & nouns & other verbs (closed_class)

# File : adjs
open(ADJS, ">$ARGV[1]") || die("Enter a filename to write adjs to");
for (my $k=0; $k<=$#adjs; $k++) {
    print (ADJS "$adjs[$k]\n");
}
close(ADJS);

# File 2: nouns
open(NOUNS, ">$ARGV[2]") || die("Enter a filename to write nouns to");
for (my $k=0; $k<=$#nouns; $k++) {
    print (NOUNS "$nouns[$k]\n");
}
close(NOUNS);

# File 3: all verbs
open(ALL_VERBS, ">$ARGV[3]") || die("Enter a filename to write all verbs to");
for (my $k=0; $k<=$#verbs; $k++) {
    print (ALL_VERBS "$verbs[$k]\n");
}
for (my $k=0; $k<=$#verbs_a_m; $k++) {
    print (ALL_VERBS "$verbs_a_m[$k]\n");
```

```
}
close(ALL_VERBS);

# File 4: open_class words (adjs, nouns, other_verbs) [Note: file handle should be OPEN_CLASS]
open(CLOSED_CLASS, ">$ARGV[4]") || die("Enter a filename to write closed class words to");
for (my $k=0; $k<=$#adjs; $k++) {
    print (CLOSED_CLASS "$adjs[$k]\n");
}
for (my $k=0; $k<=$#nouns; $k++) {
    print (CLOSED_CLASS "$nouns[$k]\n");
}
for (my $k=0; $k<=$#verbs; $k++) {
    print (CLOSED_CLASS "$verbs[$k]\n");
}
close(CLOSED_CLASS);


```

Ref: 8

Filename: Count_verb_types_from_tokens.pl

```
#!/usr/bin/perl

# Read in curriculum/sample file

open(SAMPLE, $ARGV[0]) || die("Enter a filename to open");

while (my $line = <SAMPLE>) {
    chomp($line);
    push(@words, $line);    # add each of the elements to the array @words
}   #end while
close(SAMPLE);

#############################################################################
# Convert all case to lower so comparison(eq/ne) does not take this into
# account

for (my $i=1; $i<=$#words; $i++) {
    $words[$i]=lc($words[$i]);
}

#############################################################################

#print("@unique_freq_verbs");
push (@verb_type, $words[0]);

for (my $i=1; $i<=$#words; $i++) {      # loops around @freq_verbs
    #print("\nfreq verb is: $verb_type[$i]");
    $yes_no=0;

    for (my $j=0; $j<=$#verb_type; $j++) {  # loops around @unique_freq_verbs
        #print("\nUnique Word is: $verb_type[$j]");
        if ($words[$i] eq $verb_type[$j]) {
            #print("\nValue of yes_no is: $yes_no");
            $yes_no=1;
            #print("\nValue of yes_no after compare is: $yes_no");
        }   # end if
    }   # end for @unique_freq_verbs

    if ($yes_no==0) {
        push (@verb_type, $words[$i]);
    }   # end if


#print ("\nVerb type array is: @verb_type");

}   # end for @freq_verbs

$num_of_verb_type=scalar (@verb_type);
$num_of_verb_tokens=scalar (@words);

print ("\nUnique verb types: $num_of_verb_type\n\n");
#print ("@unique_freq_verbs");

#############################################################################
# Open new file to write out results to

open(VERB_TYPES, ">$ARGV[1]") || die("Enter a filename to write out to");

print (VERB_TYPES "The verb types in file $ARGV[0] are:\n\n");
for (my $k=0; $k<=$#verb_type; $k++) {
    print (VERB_TYPES "$verb_type[$k]\n");
}
print (VERB_TYPES "\nTotal number of Verb types is: $num_of_verb_type\n");
print (VERB_TYPES "Total number of Verb tokens is: $num_of_verb_tokens");
close(VERB_TYPES);


```

Ref: 9

Filename: Count_compare_to_freq_lists_output_tokens.pl

```
#!/usr/bin/perl


# count_compare_to_freq_lists_output_tokens_types.pl
# To run: perl -w count_compare_to_freq_lists.pl
# 'sample file with data (curriculum/samples)'
# 'frequency list file' 'new file for results infreq' 'New file for results freq'

# perl -w count_compare_to_freq_lists.pl curr_corpus_verbs.txt
# leipzig_top1000de_VERBS_FROM_20.txt
# curr_basic_verb_types.txt curr_soph_verb_types.txt curr_basic_verb_tokens.txt
# curr_soph_verb_tokens.txt

# Read in the file corpus
# Read in the file freq words
# Compare files - count the number of words in the freq list against the corpora
# write out the freqs to a file (to evaluate frequencies using wordsmith tools)

# @freq_verbs == basic verb TOKENS
# @infreq_verbs == sophisticated verb TOKENS
# @unique_freq_verbs == basic verb TYPES
# @unique_infreq_verbs == sophisticated verb TYPES

# Read in curriculum/sample file

open(SAMPLE, $ARGV[0]) || die("Enter a filename to open");

while (my $line = <SAMPLE>) {
    chomp($line);
    push(@words, $line);    # add each of the elements to the array @words
}   #end while
close(SAMPLE);
print("Read in curr data file successfully\n");

# Read in frequency list file

open(FREQ, $ARGV[1]) || die("Enter a filename to open");

while (my $line = <FREQ>) {
    chomp($line);
```

```
    push(@freq, $line); # add each of the elements to the array @words
}   #end while
close(FREQ);

print("Read in freq data file successfully\n");

#################################################################
# Write to two arrays                                           ##
# Array 1: Words in curr also found in freq list -- @freq_verbs ##
# Array 2: verbs in curr not in freq list -- @infreq_verbs      ##
#################################################################

for (my $i=0; $i<=$#words; $i++) {
    # Convert all case to lower so comparison(eq/ne) does not take this into account
    $words[$i]=lc($words[$i]);

    $yes_no_freq=0;

    for (my $j=0; $j<=$#freq; $j++) {
        # Convert all case to lower so comparison(eq/ne) does not take this into account
        $freq[$i]=lc($freq[$i]);

        if ($words[$i] eq $freq[$j]) {      # if the curr word is in frequent list
            $yes_no_freq=1;     # write the word out to the array @freq_verbs
            #print("\nFound a freq verb: $words[$i]");
        }   #end if
    }   #end for freq list

    if ($yes_no_freq==0) {
        push(@infreq_verbs, $words[$i]);
    }   # end if
    else {
        push(@freq_verbs, $words[$i]);
    }
}   #end for curr/sample

$num_of_freq_verbs=scalar (@freq_verbs);
#print("\nNumber of frequent (basic) lexical tokens is: $num_of_freq_verbs");

$num_of_infreq_verbs=scalar (@infreq_verbs);
#print("\nNumber of infrequent (sophisticated) lexical tokens is: $num_of_infreq_verbs");

#######################################################
## Write out unique freq and infreq verbs           ##
## Count the number of UNIQUE freq and infreq verbs ##
#######################################################

push (@unique_freq_verbs, $freq_verbs[0]);

for (my $i=1; $i<=$#freq_verbs; $i++) {     # loops around @freq_verbs

    # Convert all case to lower so comparison(eq/ne) does not take this into account
    $freq_verbs[$i]=lc($freq_verbs[$i]);

    $yes_no=0;

    for (my $j=0; $j<=$#unique_freq_verbs; $j++) {  # loops around @unique_freq_verbs

        if ($freq_verbs[$i] eq $unique_freq_verbs[$j]) {
            $yes_no=1;
        }   # end if
    }   # end for @unique_freq_verbs

    if ($yes_no==0) {
        push (@unique_freq_verbs, $freq_verbs[$i]);
    }   # end if

}   # end for @freq_verbs

$num_of_unique_freq_verbs=scalar (@unique_freq_verbs);
print("\nNumber of unique frequent (basic) word types is: $num_of_unique_freq_verbs");

##############
##INFREQUENT##
##############
```

```
push (@unique_infreq_verbs, $infreq_verbs[0]);

for (my $i=1; $i<=$#infreq_verbs; $i++) {       # loops around @freq_verbs

    # Convert all case to lower so comparison(eq/ne) does not take this into account
    $infreq_verbs[$i]=lc($infreq_verbs[$i]);

    $yes_no=0;

    for (my $j=0; $j<=$#unique_infreq_verbs; $j++) {    # loops around @unique_freq_verbs

        if ($infreq_verbs[$i] eq $unique_infreq_verbs[$j]) {
            $yes_no=1;
        }   # end if
    }   # end for @unique_freq_verbs

    if ($yes_no==0) {
        push (@unique_infreq_verbs, $infreq_verbs[$i]);
    }   # end if

}   # end for @freq_verbs

$num_of_unique_infreq_verbs=scalar (@unique_infreq_verbs);
print("\nNumber of unique infrequent (sophisticated) word types is: $num_of_unique_infreq_verbs");

#############################################################
## Write out list of @freq_verbs, @infreq_verbs


# Open new file to write out results to :: BASIC VERB TYPES
open(FREQ_VERBS, ">$ARGV[2]") || die("Enter a filename to write out to");

print (FREQ_VERBS "\nThe frequent (basic) word types in file $ARGV[0] are:\n");
for (my $k=0; $k<=$#unique_freq_verbs; $k++) {
    print (FREQ_VERBS "$unique_freq_verbs[$k]\n");
}
print (FREQ_VERBS "Total number of basic word types is: $num_of_unique_freq_verbs\n");
close(FREQ_VERBS);



# Open new file to write out results to :: SOPHISTICATED VERB TYPES
open(INFREQ_VERBS, ">$ARGV[3]") || die("Enter a filename to write out to");

print (INFREQ_VERBS "\nThe infrequent (sophisticated) word types in file $ARGV[0] are:\n");
for (my $k=0; $k<=$#unique_infreq_verbs; $k++) {
    print (INFREQ_VERBS "$unique_infreq_verbs[$k]\n");
}
print (INFREQ_VERBS "Total number of sohisticated word types is: $num_of_unique_infreq_verbs\n");
close(INFREQ_VERBS);
```

Appendix P Automatically processing the datasets and extracting content for controlled language analysis: Perl script listing and associated code

File Ref.	File name	Description
10	Count_punc_per_sent.pl	Counts the number of punctuation marks in a file to be divided by the number of sentences in the file (indicated by sentential-final punctuation such as . , ! ? : ;).
11	Count_store_conj_per_file.pl	Counts and stores the number of conjunctions per file to be divided by the number of sentences in the file.
12	Count_store_rel_pron_per_file.pl	Counts and stores the number of relative pronouns per file to be divided by the number of sentences in the file.
13	Reading_into_array.pl	Computes the average word length in a file (to nearest integer)

Ref: 10

Filename: Count_punc_per_sent.pl

```
#!/usr/bin/perl

# count_punc_per_sent.pl

# Count all punctuation in the file . , ! ? : ; -
# Output total no. of punctuation marks in file
# This number can then be divided by the no. of sentences according to WordSmith tools

# open the file with one word per line
open(F, $ARGV[0]) || die("Enter a filename to open");

while ($line = <F>) {
    chomp($line);
    push(@words, $line);                    #split each line of the corpus at the tab and add each
                                            # of the three elements to the array @words
}   #end while
close(F);

#print ("\n @words \n");

# ARRAY @words holds the corpus: $i is the word

$num_stop=0;
$num_comma=0;
$num_colon=0;
$num_semi_colon=0;
$num_exclam=0;
$num_quest=0;
$num_hyphen=0;

open(NEW, ">$ARGV[1]") || die("Enter a filename for the new file");

#for (my $i=0; $i <= $#words ; $i+=2) {
#   #print("$words[$i]");
#   push(@just_words, $words[$i]);
#}

for (my $k=0; $k<=$#words; $k++) {

        if ($words[$k] eq '.') {        # if the token is '.'
        $num_stop+=1;                   # add one to the $num_stop counter

       }   #end if

       if ($words[$k] eq ',') {        # if the token is ','
       $num_comma+=1;                  # add one to the $num_comma counter
       }   #end if

       if ($words[$k] eq '!') {        # if the token is '!'
       $num_exclam+=1;                 # add one to the $num_exclam counter
       }   #end if

       if ($words[$k] eq '?') {        # if the token is '?'
       $num_quest+=1;                  # add one to the $num_quest counter
       }   #end if

       if ($words[$k] eq ':') {        # if the token is ':'
       $num_colon+=1;                  # add one to the $num_colon counter
       }   #end if

       if ($words[$k] eq ';') {        # if the token is ';'
       $num_semi_colon+=1;                # add one to the $num_semi_colon counter
       }   #end if

       if ($words[$k] eq '-') {        # if the token is '-'
       $num_hyphen+=1;                 # add one to the $num_hyphen counter
       }   #end if

}   # end for

$total_punc=$num_stop+$num_comma+$num_colon+$num_semi_colon+$num_exclam+$num_quest+$num_hyphen;

print (NEW "File $ARGV[0] has:\n\nMARK\t\tCOUNT\n.\t\t$num_stop\n,\t\t$num_comma\n!\t\t$num_exclam\n?\t\t$num_quest\n");
print (NEW ":\t\t$num_colon\n;\t\t$num_semi_colon\n-\t\t$num_hyphen\n");
print (NEW "\nTotal number of punctuation marks is: $total_punc");

close(NEW);
```

Ref: 11

Filename: Count_store_conj_per_file.pl

```
#!/usr/bin/perl


# count_store_conjs_per_file.pl
# To run: perl -w count_store_conjs_per_file.pl 'sample file with data' 'new file for results'


# Read in the file
# Find the conjunctions (KOUS, KOUI, KON, KOKOM)
# Count number or each type of conj
# write out the conjs to a file (to evaluate frequencies using wordsmith tools)


# Read in file

open(SAMPLE, $ARGV[0]) || die("Enter a filename to open");

while (my $line = <SAMPLE>) {
    chomp($line);
    push(@words , (split(/[\t]/, $line)));  #split each line of the corpus at the tab and add each
                                            # of the three elements to the array @words
}   #end while
close(SAMPLE);

#print("Read in data file successfully\n");
#Find the conjunctions (file is format WORD TAB POS, so CONJS can be found at all odd indexes)

for (my $i=0; $i<=$#words; $i+=2) {

    if ($words[$i+1] eq 'KON') {        # if the POS is KON
        push(@kon, $words[$i]);     # write the word out to the array @kon
        #print("Found a kon");
    }

    if ($words[$i+1] eq 'KOUI') {       # if the POS is KOUI
        push(@koui, $words[$i]);        # write the word out to the array @koui
        #print("Found a koui");
    }

    if ($words[$i+1] eq 'KOUS') {       # if the POS is KOUS
        push(@kous, $words[$i]);        # write the word out to the array @kous
```

```
        #print("Found a kous");
    }

    if ($words[$i+1] eq 'KOKOM') {      # if the POS is KOKOM
        push(@kokom, $words[$i]);       # write the word out to the array @kokom
        #print("Found a kokom");
    }

}   #end for


# Count number of each type of conjunction

$num_of_kon=scalar (@kon);
$num_of_koui=scalar (@koui);
$num_of_kous=scalar (@kous);
$num_of_kokom=scalar (@kokom);


# Open new file to write out results to
open(NUM_CONJS, ">$ARGV[1]") || die("Enter a filename to write out to");

print (NUM_CONJS "\nThe co-ordinating conjunctions in file $ARGV[0] are:\n");
for (my $k=0; $k<=$#kon; $k++) {
    print (NUM_CONJS "$kon[$k]\n");
}
print (NUM_CONJS "Total number of KONs is: $num_of_kon\n");

print (NUM_CONJS "\nThe subordinating (KOUI) conjunctions in file $ARGV[0] are:\n");
for (my $k=0; $k<=$#koui; $k++) {
    print (NUM_CONJS "$koui[$k]\n");
}
print (NUM_CONJS "Total number of KOUIs is: $num_of_koui\n");

print (NUM_CONJS "\nThe subordinating (KOUS) conjunctions in file $ARGV[0] are:\n");
for (my $k=0; $k<=$#kous; $k++) {
    print (NUM_CONJS "$kous[$k]\n");
}
print (NUM_CONJS "Total number of KOUSs is: $num_of_kous\n");


print (NUM_CONJS "\nThe subordinating (KOKOM) conjunctions are:\n");
for (my $k=0; $k<=$#kokom; $k++) {
    print (NUM_CONJS "$kokom[$k]\n");
}
print (NUM_CONJS "Total number of KOKOMs is: $num_of_kokom\n");

$total_num_of_subord=$num_of_koui+$num_of_kous+$num_of_kokom;

print (NUM_CONJS "OVERALL RESULTS:\nTotal number of co-ordinating conjunctions is: $num_of_kon\n");
print (NUM_CONJS "Total number of subordinating conjunctions is: $total_num_of_subord\n");

close(NUM_CONJS);

```

Ref: 12

Filename: Count_store_rel_pron_per_file.pl

```
#!/usr/bin/perl

# count_store_rel_pron_per_file.pl
# To run: perl -w count_store_rel_pron_per_file.pl 'sample file with data[word+POS]' 'new file for

# Read in the file
# Find the relative pronouns (PRELS, PRELAT)
# Count number or each type of relative pronoun
# write out the relative pronouns to a file (to evaluate frequencies using wordsmith tools)

# Read in file

open(SAMPLE, $ARGV[0]) || die("Enter a filename to open");

while (my $line = <SAMPLE>) {
    chomp($line);
    push(@words , (split(/[\t]/, $line)));  #split each line of the corpus at the tab and add each
                                            # of the three elements to the array @words
}   #end while
close(SAMPLE);

#print("Read in data file successfully\n");
#Find the conjunctions (file is format WORD TAB POS, so CONJS can be found at all odd indexes)

for (my $i=0; $i<=$#words; $i+=2) {

    if ($words[$i+1] eq 'PRELS') {      # if the POS is PRELS
        push(@prels, $words[$i]);       # write the word out to the array @prels
        #print("Found a prels");
    }

    if ($words[$i+1] eq 'PRELAT') {     # if the POS is PRELAT
        push(@prelat, $words[$i]);      # write the word out to the array @prelat
        #print("Found a prelat");
    }

}   #end for


# Count number of each type of conjunction


$num_of_prels=scalar (@prels);
$num_of_prelat=scalar (@prelat);


# Open new file to write out results to
open(NUM_REL_PRON, ">$ARGV[1]") || die("Enter a filename to write out to");

print (NUM_REL_PRON "\nThe number of attributive relative pronouns in file $ARGV[0] are:\n");
for (my $k=0; $k<=$#prels; $k++) {
    print (NUM_REL_PRON "$prels[$k]\n");
}
print (NUM_REL_PRON "Total number of PRELs is: $num_of_prels\n");

print (NUM_REL_PRON "\nThe substitutive relative pronouns in file $ARGV[0] are:\n");
for (my $k=0; $k<=$#prelat; $k++) {
    print (NUM_REL_PRON "$prelat[$k]\n");
}
print (NUM_REL_PRON "Total number of PRELATs is: $num_of_prelat\n");

$total_num_of_rel_prons=$num_of_prels+$num_of_prelat;

print (NUM_REL_PRON "OVERALL RESULT:\nTotal number of relative pronouns is: $total_num_of_rel_prons\n");
#print (NUM_CONJS "Total number of subordinating conjunctions is: $total_num_of_subord\n");

close(NUM_REL_PRON);

```

Ref: 13

Filename: Reading_into_array.pl

```
#!/usr/bin/perl

#print "Hello, World...\n";

open(F, $ARGV[0]) || die("Enter a filename");

while ($line = <F>) {
    chomp($line);
    push(@words , (split(/[\t]/, $line)));  #split each line of the corpus at the tab and add each
                                            # of the three elements to the array @words
}   #end while
close(F);

print ("\n @words \n");

# $i is the word, $i+1 is the POS, $i+2 is the lemma

################################
# Compute average word length #
################################
$sum_of_word_lengths=0;
$num_of_words=0;
$average_length=0;

for (my $i=0; $i <= $#words; $i+3) {
    $sum_of_word_lengths += length($words[i]);
    $num_of_words++;
}

$average_length = $sum_of_word_lengths / $num_of_words;
print("$average_length");

#only returns full numbers - no floats so a little inaccurate.

###################################
# Compute average sentence length #
###################################

```

Appendix Q Full data set from computational analysis of content: Phase 1

Test Data	#Tokens	#Types	Type/ Token Ratio (TTR)	Stan-ised TTR (STTR)	Average Word Length (chars)	Average Sent. Length (words)	# punc. marks/sent. (inc. final .)	# co-ord. conjs	# subord. conjs
Raw Content	2,887	909	32.23	36.25	4.84	3.18 (skewed)	1.15	3	1 (eg. in 4 sens.)
Con_sample1	3,686	1,606	44.61	56.17	6.04	17.31	2	92	53
Con_sample2	3,646	1,527	42.77	52.93	6.21	15.00	1.65	93	45
Con_sample3	3,569	1,555	44.29	56.73	6.08	16.96	2.02	69	48
Con_sample4	3,642	1,570	44.11	54.90	6.01	15.61	1.73	81	56
Con_sample5	3,660	1,562	43.81	54.83	5.98	15.84	1.75	80	34
Con_sample6	3,625	1,593	44.70	54.40	6.23	19.69	2.11	87	39
Con_sample7	3,627	1,599	45.13	56.80	6.15	19.15	2.07	111	50
Con_sample8	3,544	1,572	44.89	53.37	6.39	22.59	2.59	122	66
Con_sample9	3,542	1,528	43.77	52.89	5.94	18.87	2.22	144	49
Con_sample10	3,566	1,623	46.48	56.80	6.15	17.46	2.16	103	64
AVERAGE	**3,610.7**	**1,573.5**	**44.46**	**54.98**	**6.12**	**17.86**	**2.03**	**98.2**	**50.4**
Ran_sample1	3,623	1,701	47.81	58.67	6.03	---	---	94	54

Ran_sample2	3,601	1,716	48.47	58.93	6.23	---	---	92	45
Ran_sample3	3,594	1,767	49.93	59.70	6.27	---	---	86	44
Ran_sample4	3,600	1,775	50.03	60.50	6.32	---	---	71	47
Ran_sample5	3,589	1,717	48.75	59.63	6.12	---	---	93	44
Ran_sample6	3,582	1,740	49.38	60.27	6.21	---	---	87	48
Ran_sample7	3,595	1,752	49.49	60.00	6.07	---	---	79	59
Ran_sample8	3,590	1,708	48.32	59.17	6.10	---	---	78	53
Ran_sample9	3,616	1,750	49.20	60.20	6.28	---	---	90	45
Ran_sample10	3,621	1,754	49.30	59.77	6.08	---	---	95	39
AVERAGE	**3,601.1**	**1,738**	**49.07**	**59.68**	**6.17**	**---**	**---**	**86.5**	**47.8**
PWD TXTBK	**3,239**	**540**	**18.11**	**26.10**	**4.58**	**5.14 (skewed)**	**1.32**	**92**	**0**

Test Data	**#Relative Pronouns** PRELS| PRELAT| TOT		
Raw Content	0	0	0
Con_sample1	20	4	24
Con_sample2	25	1	26
Con_sample3	24	1	25
Con_sample4	28	1	29
Con_sample5	27	0	27
Con_sample6	19	2	21
Con_sample7	18	2	20
Con_sample8	31	1	32
Con_sample9	33	2	35
Con_sample10	27	0	27
AVERAGE	25.2	1.4	26.6
Ran_sample1	23	3	26
Ran_sample2	29	1	30
Ran_sample3	25	0	25

Ran_sample4	28	4	32					
Ran_sample5	24	3	27					
Ran_sample6	25	2	27					
Ran_sample7	32	2	34					
Ran_sample8	30	0	30					
Ran_sample9	27	5	32					
Ran_sample10	30	1	31					
AVERAGE	27.3	2.1	29.4					
PWD TXTBK	0	0	0					

Test Data	#Tokens	#Types	#Lexical word tokens	#Lexical word types	#Verb tokens	#Verb types	#Noun tokens	#Noun types	#Adj. tokens	#Adj. types
Raw Content.	**2,887**	**909**	**1124**	**667**	**433**	**115**	**737**	**444**	**191**	**127**
Con1	3,686	1,606	1785	1337	486	319	1131	801	335	266
Con2	3,646	1,527	1871	1286	513	304	1219	784	340	242
Con3	3,569	1,555	1782	1271	494	329	1114	727	328	255

Con4	3,642	1,570	1881	1325	500	290	1246	839	326	240
Con5	3,660	1,562	1832	1284	475	284	1167	773	369	273
Con6	3,625	1,593	1782	1328	510	310	1080	755	394	313
Con7	3,627	1,599	1702	1291	462	311	1035	752	364	280
Con8	3,544	1,572	1712	1282	433	272	1014	725	412	329
Con9	3,542	1,528	1683	1260	476	309	1056	752	326	248
Con10	3,566	1,623	1731	1344	496	334	1095	807	315	255
AV.	**3,610.7**	**1,573.5**	**1776.1**	**1300.8**	**484.5**	**306.2**	**1115.7**	**771.5**	**350.9**	**270.1**
Ran1	3,623	1,701	1754	1407	469	284	1153	913	326	259
Ran2	3,601	1,716	1834	1441	508	294	1152	902	359	285
Ran3	3,594	1,767	1842	1467	483	316	1175	924	335	271
Ran4	3,600	1,775	1870	1500	500	313	1216	959	334	278
Ran5	3,589	1,717	1771	1440	480	307	1156	911	317	268
Ran6	3,582	1,740	1832	1470	482	305	1181	929	347	284
Ran7	3,595	1,752	1837	1456	477	304	1191	936	330	262
Ran8	3,590	1,708	1730	1418	484	288	1109	893	328	279
Ran9	3,616	1,750	1807	1475	502	317	1148	927	338	274
Ran10	3,621	1,754	1766	1455	477	290	1143	923	340	294
AV.	**3,601.1**	**1,738**	**1804.3**	**1452.9**	**486.2**	**301.8**	**1162.4**	**921.7**	**335.4**	**275.4**

PWD	**3,239**	**540**	**1,429**	**403**	**467**	**77**	**963**	**306**	**95**	**41**

Lexical words: open-class words → noun, adjectives and verbs (except auxiliaries and modals)

[Note: mistake in labeling the files: open-class words are actually labeled as closed-class]

Sophisticated types/tokens: those not found in [100] most frequently used word types/verbs/nouns

Basic types/tokens: those found in [100] most freq used word types/verbs/nouns

Measures	**Raw Content**	**Con. Sample Av.**	**Ran. Sample Ave.**	**PWD Textbook**
Word Variation-1 **WT/W** Tot. types/total tokens	0.31	0.44	0.48	0.17
Word Variation-2 **WT/√2W**	11.96	37.03	40.95	6.71
Word Variation-3 **WT(WT)/W**	286.21	685.71	838.81	90.03
Verb Variation-1 **VT/V** Verb types/verb tokens	0.27	0.63	0.62	0.16
Verb Variation-2 **VT/LW** Verb types/lexical word tokens	0.10	0.17	0.17	0.05

Lexical Variation **LWT/LW** Lexical word types/lexical word tokens	0.59	0.73	0.81	0.28
Noun Variation **NT/LW** Noun types/lexical word tokens	0.40	0.43	0.51	0.21
Adjective Variation **AdjT/LW** Adj types/lexical word tokens	0.11	0.15	0.15	0.03
Lexical Density **LW/W** Lexical word tokens/total no. tokens	0.39	0.49	0.50	0.44

Test Data	#Soph. verb types (20 freq verbs)	# Soph. Verb types (100 freq verbs)	# Soph. Verb types (200 freq verbs)	#Soph. lexical word tokens (1000 freq words)	#Soph. lexical word tokens (2000 freq words)	#Soph. word types (1000 freq words)	#Basic word types (1000 freq words)	#Soph. word types (2000 freq words)	#Basic word types (2000 freq words)
Raw Content.	**107**	**93**	**85**	**921**	**880**	**773**	**192**	**721**	**250**
Con1	301	254	231	1332	1255	1256	428	1135	553
Con2	285	248	225	1423	1330	1205	388	1089	510
Con3	310	269	243	1347	1253	1169	453	1039	593
Con4	272	236	213	1384	1276	1227	441	1095	553
Con5	265	223	201	1311	1200	1193	450	1055	594
Con6	291	243	218	1353	1252	1227	437	1088	580
Con7	292	250	225	1319	1213	1238	426	1096	572
Con8	253	208	186	1318	1224	1199	406	1081	528
Con9	290	245	225	1348	1244	1168	407	1013	568
Con10	316	268	230	1369	1307	1235	455	1122	569
AV.	**287.5**	**244.4**	**219.7**	**1350.4**	**1255.4**	**1211.7**	**429.1**	**1081.3**	**562**

Ran1	265	223	197	1377	1292	1292	455	1170	587
Ran2	275	232	208	1425	1339	1332	435	1188	583
Ran3	298	256	229	1445	1337	1350	456	1200	611
Ran4	295	242	218	1449	1342	1343	463	1192	619
Ran5	288	247	217	1403	1319	1332	437	1187	586
Ran6	286	239	209	1436	1353	1329	450	1201	587
Ran7	286	242	213	1445	1368	1352	445	1207	596
Ran8	269	226	192	1360	1276	1301	452	1161	598
Ran9	298	254	227	1414	1322	1359	434	1214	583
Ran10	271	229	203	1396	1308	1340	457	1199	605
AV.	**283.1**	**239**	**211.3**	**1415**	**1325.6**	**1333**	**448.4**	**1191.9**	**595.5**
PWD	**75**	**61**	**55**	**352**	**325**	**491**	**160**	**440**	**221**

Measures	**Raw Content**			**Consecutive Sample Av.**			**Random Sample Average**		
Comparison with the X most frequent	**20 freq**	**100 freq**	**200 freq**	**20 freq**	**100 freq**	**200 freq**	**20 freq**	**100 freq**	**200 freq**
Verb Sophistication **SVT/V** Tot. soph verb types/total verb tokens	0.25	0.21	0.2	0.6	0.5	0.45	0.6	0.5	0.44
Verb Sophistication 2 **SVT^2/V** Tot. soph verb types squared/total verb tokens	26.44	19.97	16.69	170.6	123.28	99.62	168.02	119.75	93.60

Verb Sophistication 3 **SVT/√2V** Tot. soph verb types/sq root 2*total verb tokens	3.64	3.16	2.89	12.05	10.25	9.21	9.16	9.49	6.84
	1000 freq		**2000 freq**	**1000 freq**		**2000 freq**	**1000 freq**		**2000 freq**
Lexical Sophistication 1 **SLW/LW** Soph lex word tokens/lex word tokens	0.82		0.78	0.76		0.71	0.8		0.75
Lexical Sophistication 2 **SWT/WT** Soph word types/word types	0.85		0.79	0.69		0.27	0.76		0.68
Lexical Basicness **BWT/WT** Basic word types/word types	0.21		0.28	0.27		0.36	0.26		0.34

Measures	**PWD Textbook**		
	20 freq	**100 freq**	**200 freq**
Verb Sophistication **SVT/V** Tot. soph verb types/total verb tokens	0.16	0.13	0.12
Verb Sophistication 2 **SVT^2/V** Tot. soph verb types squared/total verb tokens	12.04	7.97	6.48
Verb Sophistication 3 **SVT/√2V** Tot. soph verb types/sq root 2*total verb tokens	2.45	2.00	1.80

	1000 freq	**2000 freq**
Lexical Sophistication 1 **SLW/LW** Soph lex word tokens/lex word tokens	0.25	0.23
Lexical Sophistication 2 **SWT/WT** Soph word types/word types	0.91	0.81
Lexical Basicness **BWT/WT** Basic word types/word types	0.30	0.41

Appendix R Full data set from computational analysis of content: Phase 2

Test Data	#Tokens	#Types	Type/ Token Ratio (TTR)	Stan-ised TTR	Average Word Length (chars)	Average Sent. Length (words)	# punc. marks/sent. (inc. final .)	# co-ord. conjs	# subord. conjs
Raw Content	**1,880**	**467**	**25.07**	**27.40**	**4.51**	**4.05**	**1.04**	**2**	**4 (for 1 kokom als)**
Con_sample1	2,280	1,111	49.69	57.35	6.07	18.03	2.15	63	36
Con_sample2	2,253	1,043	47.15	54.10	6.17	14.93	1.68	52	28
Con_sample3	2,228	1,057	48.35	55.90	6.02	15.84	1.96	48	31
Con_sample4	2,284	1,076	48.19	54.80	6.14	15.51	1.61	48	17
Con_sample5	2,282	1,054	47.54	54.15	5.85	16.30	1.80	56	21
Con_sample6	2,240	1,045	47.29	52.75	6.32	19.91	2.10	75	23
Con_sample7	2,273	1,095	49.32	56.80	6.16	19.65	2.19	87	43
Con_sample8	2,220	1,034	47.32	52.30	5.88	19.00	2.23	55	26
Con_sample9	2,214	1,054	48.30	54.20	5.98	20.17	2.27	60	39
Con_sample10	2,253	1,072	48.77	54.40	6.04	17.58	2.02	47	22
AVERAGE	**2,252.7**	**1,064.1**	**48.19**	**54.68**	**6.06**	**17.69**	**2**	**59.1**	**28.6**
Ran_sample1	2,256	1,130	50.95	57.05	6.09	---	---	60	31

Ran_sample2	2,267	1,147	51.60	57.55	6.27	---	---	63	29
Ran_sample3	2,222	1,145	52.43	58.00	6.32	---	---	50	26
Ran_sample4	2,230	1,158	52.68	59.10	6.32	---	---	42	29
Ran_sample5	2,236	1,107	50.20	57.80	6.05	---	---	68	32
Ran_sample6	2,232	1,156	52.62	59.25	6.27	---	---	51	35
Ran_sample7	2,255	1,130	50.92	57.75	6.07	---	---	56	37
Ran_sample8	2,237	1,131	51.22	57.75	6.12	---	---	48	30
Ran_sample9	2,224	1,160	53.04	60.20	6.31	---	---	55	23
Ran_sample10	2,245	1,144	51.65	58.65	6.04	---	---	65	21
AVERAGE	**2,240.4**	**1,140.8**	**51.73**	**58.31**	**6.19**	**---**	**---**	**55.8**	**29.3**
PWD TXTBK	**1,839**	**344**	**18.88**	**20.60**	**4.75**	**3.93**	**1.14**	**78**	**0**

Test Data	**#Relative Pronouns** PRELS| PRELAT| TOT							
Raw Content	0	0	0					
Con_sample1	9	3	12					
Con_sample2	17	0	17					
Con_sample3	17	1	18					
Con_sample4	13	0	13					
Con_sample5	18	0	18					
Con_sample6	10	2	12					
Con_sample7	15	1	16					
Con_sample8	18	1	19					
Con_sample9	21	1	22					
Con_sample10	13	0	13					
AVERAGE	15.1	0.9	16					
Ran_sample1	14	2	16					
Ran_sample2	19	0	19					
Ran_sample3	12	0	12					

Ran_sample4	19	0	19					
Ran_sample5	16	3	19					
Ran_sample6	13	2	15					
Ran_sample7	19	1	20					
Ran_sample8	21	0	21					
Ran_sample9	15	3	18					
Ran_sample10	22	1	23					
AVERAGE	17	1.2	18.2					
PWD TXTBK	0	0	0					

Test Data	#Tokens	#Types	#Lexical word tokens	#Lexical word types	#Verb tokens	#Verb types	#Noun tokens	#Noun types	#Adj. tokens	#Adj. types
Raw Content	**1,880**	**467**	**673**	**341**	**422**	**106**	**369**	**182**	**119**	**70**
Con1	2,280	1,111	1050	875	311	221	633	511	214	181

Con2	2,253	1,043	1081	841	319	208	685	497	203	174
Con3	2,228	1,057	1075	832	299	217	660	472	205	176
Con4	2,284	1,076	1124	885	305	193	754	572	180	155
Con5	2,282	1,054	1094	835	291	179	687	508	228	179
Con6	2,240	1,045	1060	829	333	207	624	457	250	208
Con7	2,273	1,095	1028	841	296	218	595	465	238	198
Con8	2,220	1,034	1044	803	298	205	647	472	206	167
Con9	2,214	1,054	1015	835	288	201	631	498	202	175
Con10	2,253	1,072	1092	870	295	214	714	537	185	157
AV.	**2,252.7**	**1,064.1**	**1066.3**	**844.6**	**303.5**	**206.3**	**663**	**498.9**	**211.1**	**177**
Ran1	2,256	1,130	1078	885	288	184	718	576	191	164
Ran2	2,267	1,147	1135	924	330	205	714	575	206	179
Ran3	2,222	1,145	1120	901	290	193	704	561	216	179
Ran4	2,230	1,158	1132	928	314	203	716	576	212	185
Ran5	2,236	1,107	1053	872	304	201	678	543	191	167
Ran6	2,232	1,156	1117	925	303	197	725	586	203	174
Ran7	2,255	1,130	1101	892	302	203	702	559	198	164
Ran8	2,237	1,131	1076	896	287	177	679	569	210	185
Ran9	2,224	1,160	1106	925	314	212	694	569	215	181

Ran10	2,245	1,144	1075	896	283	177	697	567	214	190
AV.	**2,240.4**	**1,140.8**	**1099.3**	**904.4**	**301.5**	**195.2**	**702.7**	**568.1**	**205.6**	**176.8**
PWD	**1,839**	**344**	**893**	**268**	**341**	**59**	**576**	**196**	**50**	**28**

Lexical words: open-class words → noun, adjectives and verbs (except auxiliaries and modals)

[Note: mistake in labeling the files: open-class words are actually labeled as closed-class]

Sophisticated types/tokens: those not found in [100] most frequently used word types/verbs/nouns

Basic types/tokens: those found in [100] most freq used word types/verbs/nouns

Measures	**Raw Content**	**Consecutive Sample Av.**	**Random Sample Av.**	**PWD Textbook**
Word Variation-1 **WT/W** Tot. types/total tokens	0.40	0.47	0.51	0.19
Word Variation-2 **WT/√2W**	7.62	15.83	17.00	5.67
Word Variation-3 **WT(WT)/W**	116.00	2252.7	580.89	64.35
Verb Variation-1 **VT/V** Verb types/verb tokens	0.25	0.68	0.65	0.17
Verb Variation-2	0.16	0.19	0.18	0.07

VT/LW Verb types/lexical word tokens				
Lexical Variation **LWT/LW** Lexical word types/lexical word tokens	0.51	0.79	0.82	0.30
Noun Variation **NT/LW** Noun types/lexical word tokens	0.27	0.47	0.51	0.22
Adjective Variation **AdjT/LW** Adj types/lexical word tokens	0.10	0.17	0.16	0.03
Lexical Density **LW/W** Lex word tokens/total no. tokens	0.36	0.47	0.49	0.49

Test Data	#Soph. verb types (20 freq verbs)	# Soph. Verb types (100 freq verbs)	# Soph. Verb types (200 freq verbs)	#Soph. lexical word tokens (1000 freq)	#Soph. lexical word tokens (2000 freq)	#Soph. word types (1000 freq words)	#Basic word types (1000 freq words)	#Soph. word types (2000 freq words)	#Basic word types (2000 freq words)

Raw content.	**98**	**84**	**77**	**552**	**519**	**338**	**147**	**302**	**184**
Con1	208	172	156	855	809	810	343	739	416
Con2	190	161	145	925	883	782	298	725	357
Con3	198	170	154	877	830	762	347	691	425
Con4	178	151	134	921	860	822	299	742	382
Con5	161	134	119	836	781	758	350	681	430
Con6	188	152	137	852	800	767	328	689	409
Con7	200	169	152	847	782	810	332	723	420
Con8	186	146	128	833	786	746	319	680	388
Con9	184	151	138	865	814	780	311	700	395
Con10	197	166	138	908	869	800	237	732	393
AV.	**189**	**157.2**	**140.1**	**871.9**	**821.4**	**783.7**	**316.4**	**710.2**	**401.5**
Ran1	165	134	113	901	855	807	351	740	422
Ran2	187	152	134	939	894	858	324	778	407
Ran3	177	146	129	943	889	843	331	756	421
Ran4	187	147	134	942	892	848	337	769	416
Ran5	183	154	131	904	856	804	325	731	400

Ran6	179	144	125	951	919	852	326	791	388
Ran7	185	154	135	929	890	832	325	752	408
Ran8	158	131	110	915	868	837	325	757	407
Ran9	194	161	142	928	881	863	330	786	409
Ran10	159	129	113	894	848	823	339	748	416
AV.	**177.4**	**145.2**	**126.6**	**924.6**	**879.2**	**836.7**	**331.3**	**760.8**	**409.4**
PWD	**58**	**47**	**42**	**686**	**566**	**251**	**125**	**218**	**168**

Measures	**Raw Content**			**Consecutive Sample Av.**			**Random Sample Average**		
	20 freq	**100 freq**	**200 freq**	**20 freq**	**100 freq**	**200 freq**	**20 freq**	**100 freq**	**200 freq**
Verb Sophistication **SVT/V** Tot. soph verb types/total verb tokens	0.23	0.20	0.18	0.62	0.52	0.46	0.59	0.48	0.42
Verb Sophistication 2 **SVT^2/V** Tot. soph verb types squared/total verb tokens	22.76	16.72	14.05	117.70	81.42	64.67	104.38	69.93	53.16

Verb Sophistication 3 **SVT/√2V** Tot. soph verb types/sq root 2*total verb tokens	3.37	2.89	2.65	7.67	6.38	5.69	7.22	5.91	5.16

	1000 freq	**2000 freq**	**1000 freq**	**2000 freq**	**1000 freq**	**2000 freq**
Lexical Sophistication 1 **SLW/LW** Soph lex word tokens/lex word tokens	0.82	0.77	0.82	0.77	0.84	0.80
Lexical Sophistication 2 **SWT/WT** Soph word types/word types	0.72	0.65	0.74	0.67	0.73	0.67
Lexical Basicness **BWT/WT** Basic word types/word types	0.31	0.39	0.30	0.38	0.29	0.36

Measures	**PWD Textbook**		
	20 freq	**100 freq**	**200 freq**

Verb Sophistication **SVT/V** Tot. soph verb types/total verb tokens	0.17	0.14	0.12
Verb Sophistication 2 **SVT^2/V** Tot. soph verb types squared/total verb tokens	9.87	6.48	5.17
Verb Sophistication 3 **SVT/√2V** Tot. soph verb types/sq root 2*total verb tokens	2.22	1.80	1.60

	1000 freq	**2000 freq**
Lexical Sophistication 1 **SLW/LW** Soph lex word tokens/lex word tokens	0.77	0.63
Lexical Sophistication 2 **SWT/WT** Soph word types/word types	0.73	0.63

Lexical Basicness **BWT/WT** Basic word types/word types	0.37	0.49

Appendix S Extract from the XML linguistic content database: Section One: Greetings

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE database SYSTEM "XML_DTD.dtd">

<database>
<database_desc>German modern language for primary school children in Ireland. PSC, Modern language curriculum documents and LSs Linkage, Integration, Language Awareness and Error Analysis</database_desc>

<section>
<name>Section 1: General Introductions</name>
<subsection>
	<subsection_name>1. Greetings</subsection_name>
	<multi_lang_sent>
		<sent>
			<sent_lang>German</sent_lang>
			<word pos="ITJ">Hallo</word>
			<word pos="FULL">!</word>
		</sent>
		<sent>
			<sent_lang>English</sent_lang>
			<full_sent>Hello!</full_sent>
		</sent>
		<sent>
			<sent_lang>Irish</sent_lang>
			<full_sent>Dia duit!</full_sent>
		</sent>
		<mm_file>
			<mm_type>image</mm_type>
			<mm_name>images/hello.gif</mm_name>
		</mm_file>
	</multi_lang_sent>
	<multi_lang_sent>
		<sent>
			<sent_lang>German</sent_lang>
			<word pos="ADJA">Guten</word>
			<word pos="NN">Morgen</word>
			<word pos="FULL">.</word>
		</sent>
		<sent>
			<sent_lang>English</sent_lang>
			<full_sent>Good morning.</full_sent>
		</sent>
		<sent>
			<sent_lang>Irish</sent_lang>
			<full_sent>Dia duit ar maidin.</full_sent>
		</sent>
	</multi_lang_sent>

	<multi_lang_sent>
		<sent>
			<sent_lang>German</sent_lang>
```

```
                        <word pos="ADJA">Guten</word>
                        <word pos="NN">Tag</word>
                        <word pos="FULL">.</word>
                </sent>
                <sent>
                        <sent_lang>English</sent_lang>
                        <full_sent>Good day.</full_sent>
                </sent>
                <sent>
                        <sent_lang>Irish</sent_lang>
                        <full_sent>Dia duit.</full_sent>
                </sent>
        </multi_lang_sent>
        <multi_lang_sent>
                <sent>
                        <sent_lang>German</sent_lang>
                        <word pos="ADJA">Guten</word>
                        <word pos="NN">Abend</word>
                        <word pos="FULL">.</word>
                </sent>
                <sent>
                        <sent_lang>English</sent_lang>
                        <full_sent>Good evening.</full_sent>
                </sent>
                <sent>
                        <sent_lang>Irish</sent_lang>
                        <full_sent>Dia duit.</full_sent>
                </sent>
        </multi_lang_sent>
        <multi_lang_sent>
                <sent>
                        <sent_lang>German</sent_lang>
                        <word pos="ADJA">Gute</word>
                        <word pos="NN">Nacht</word>
                        <word pos="FULL">.</word>
                </sent>
                <sent>
                        <sent_lang>English</sent_lang>
                        <full_sent>Good night.</full_sent>
                </sent>
                <sent>
                        <sent_lang>Irish</sent_lang>
                        <full_sent>OU+00EDche mhaith.</full_sent>
                </sent>
</multi_lang_sent>
<multi_lang_sent>
                <sent>
                        <sent_lang>German</sent_lang>
                        <word pos="VVIMP">Schlaf</word>
                        <word pos="ADV">gut</word>
                        <word pos="FULL">.</word>
                </sent>

                <sent>
                        <sent_lang>English</sent_lang>
                        <full_sent>Sleep well.</full_sent>
                </sent>

                <sent>
                        <sent_lang>Irish</sent_lang>
                        <full_sent>Codladh sU+00E1mh.</full_sent>
```

```
                </sent>
</multi_lang_sent>

<multi_lang_sent>
                <sent>
                        <sent_lang>German</sent_lang>
                        <word pos="APPR">Auf</word>
                        <word pos="NN">Wiedersehen</word>
                        <word pos="FULL">.</word>
                </sent>

                <sent>
                        <sent_lang>English</sent_lang>
                        <full_sent>Goodbye.</full_sent>
                </sent>

                <sent>
                        <sent_lang>Irish</sent_lang>
                        <full_sent>SlU+00E1n.</full_sent>
                </sent>
</multi_lang_sent>

<multi_lang_sent>
                <sent>
                        <sent_lang>German</sent_lang>
                        <word pos="ITJ">TschU+00FCs</word>
                        <word pos="FULL">.</word>
                </sent>

                <sent>
                        <sent_lang>English</sent_lang>
                        <full_sent>Goodbye.</full_sent>
                </sent>

                <sent>
                        <sent_lang>Irish</sent_lang>
                        <full_sent>SlU+00E1n.</full_sent>
                </sent>
</multi_lang_sent>

<multi_lang_sent>
                <sent>
                        <sent_lang>German</sent_lang>
                        <word pos="APPR">Bis</word>
                        <word pos="ADV">bald</word>
                        <word pos="FULL">.</word>
                </sent>

                <sent>
                        <sent_lang>English</sent_lang>
                        <full_sent>See you soon.</full_sent>
                </sent>

                <sent>
                        <sent_lang>Irish</sent_lang>
                        <full_sent>Feicfidh mU+00E9 go luath tU+00FA.</full_sent>
                </sent>
</multi_lang_sent>

<multi_lang_sent>
                <sent>
```

```
                        <sent_lang>German</sent_lang>
                        <word pos="APPR">Bis</word>
                        <word pos="ADV">morgen</word>
                        <word pos="FULL">.</word>
                </sent>

                <sent>
                        <sent_lang>English</sent_lang>
                        <full_sent>See you in the morning.</full_sent>
                </sent>

                <sent>
                        <sent_lang>Irish</sent_lang>
                        <full_sent>Feicfidh mU+00E9 ar maidin tU+00FA.</full_sent>
                </sent>
</multi_lang_sent>

<multi_lang_sent>
                <sent>
                        <sent_lang>German</sent_lang>
                        <word pos="APPR">Bis</word>
                        <word pos="NN">Donnerstag</word>
                        <word pos="FULL">.</word>
                </sent>

                <sent>
                        <sent_lang>English</sent_lang>
                        <full_sent>See you on Thursday.</full_sent>
                </sent>

                <sent>
                        <sent_lang>Irish</sent_lang>
                        <full_sent>Feicfidh mU+00E9 dU+00E9 DU+00E9ardaoin
tU+00FA.</full_sent>
                </sent>
</multi_lang_sent>

</subsection>

<subsection>
<subsection_name>How are you?</subsection_name>

<multi_lang_sent>
                <sent>
                        <sent_lang>German</sent_lang>
                        <word pos="PWAV">Wie</word>
                        <word pos="VVFIN">geht</word>
                        <word pos="PPER">'s</word>
                        <word pos="FULL">?</word>
                </sent>

                <sent>
                        <sent_lang>English</sent_lang>
                        <full_sent>How are you?</full_sent>
                </sent>

                <sent>
                        <sent_lang>Irish</sent_lang>
                        <full_sent>Conas atU+00E1 tU+00FA?</full_sent>
                </sent>
</multi_lang_sent>
```

```
<multi_lang_sent>
		<sent>
			<sent_lang>German</sent_lang>
			<word pos="ADJD">Gut</word>
			<word pos="COMMA">,</word>
			<word pos="PTKANT">danke</word>
			<word pos="FULL">.</word>
		</sent>

		<sent>
			<sent_lang>English</sent_lang>
			<full_sent>Good, thank you.</full_sent>
		</sent>

		<sent>
			<sent_lang>Irish</sent_lang>
			<full_sent>Go maith, go raibh maith agat.</full_sent>
		</sent>
</multi_lang_sent>

<multi_lang_sent>
		<sent>
			<sent_lang>German</sent_lang>
			<word pos="ADV">Sehr</word>
			<word pos="ADJD">gut</word>
			<word pos="COMMA">,</word>
			<word pos="PTKANT">danke</word>
			<word pos="FULL">.</word>
		</sent>

		<sent>
			<sent_lang>English</sent_lang>
			<full_sent>Very good, thank you.</full_sent>
		</sent>

		<sent>
			<sent_lang>Irish</sent_lang>
			<full_sent>An-mhaith, go raibh maith agat.</full_sent>
		</sent>
</multi_lang_sent>

<multi_lang_sent>
		<sent>
			<sent_lang>German</sent_lang>
			<word pos="PTKNEG">Nicht</word>
			<word pos="ADV">so</word>
			<word pos="ADJD">gut</word>
			<word pos="FULL">.</word>
		</sent>

		<sent>
			<sent_lang>English</sent_lang>
			<full_sent>Not so good.</full_sent>
		</sent>

		<sent>
			<sent_lang>Irish</sent_lang>
			<full_sent>NU+00EDlim rU+00E9mhaith.</full_sent>
		</sent>
</multi_lang_sent>
```

```
<multi_lang_sent>
		<sent>
			<sent_lang>German</sent_lang>
			<word pos="ADJD">Schlecht</word>
			<word pos="FULL">.</word>
		</sent>

		<sent>
			<sent_lang>English</sent_lang>
			<full_sent>I'm terrible.</full_sent>
		</sent>

		<sent>
			<sent_lang>Irish</sent_lang>
			<full_sent>TU+00E1im go holc.</full_sent>
		</sent>
</multi_lang_sent>

<multi_lang_sent>
		<sent>
			<sent_lang>German</sent_lang>
			<word pos="PPER">Mir</word>
			<word pos="VVFIN">geht</word>
			<word pos="PPER">'s</word>
			<word pos="ADJD">gut</word>
			<word pos="FULL">.</word>
		</sent>

		<sent>
			<sent_lang>English</sent_lang>
			<full_sent>I'm good.</full_sent>
		</sent>

		<sent>
			<sent_lang>Irish</sent_lang>
			<full_sent>TU+00E1im go brea.</full_sent>
		</sent>
</multi_lang_sent>

<multi_lang_sent>
		<sent>
			<sent_lang>German</sent_lang>
			<word pos="PPER">Mir</word>
			<word pos="VVFIN">geht</word>
			<word pos="PPER">'s</word>
			<word pos="ADV">sehr</word>
			<word pos="ADJD">gut</word>
			<word pos="FULL">.</word>
		</sent>

		<sent>
			<sent_lang>English</sent_lang>
			<full_sent>I'm really well.</full_sent>
		</sent>

		<sent>
			<sent_lang>Irish</sent_lang>
			<full_sent>TU+00E1im an-mhaith.</full_sent>
		</sent>
</multi_lang_sent>
```

```
</subsection>


<subsection>
<subsection_name>What is your name?</subsection_name>

<multi_lang_sent>
		<sent>
			<sent_lang>German</sent_lang>
			<word pos="PWAV">Wie</word>
			<word pos="VVFIN">heiU+00DFt</word>
			<word pos="PPER">du</word>
			<word pos="FULL">?</word>
		</sent>

		<sent>
			<sent_lang>English</sent_lang>
			<full_sent>What is your name?</full_sent>
		</sent>

		<sent>
			<sent_lang>Irish</sent_lang>
			<full_sent>Cad is ainm duit?</full_sent>
		</sent>
</multi_lang_sent>

<multi_lang_sent>
		<sent>
			<sent_lang>German</sent_lang>
			<word pos="PPER">Ich</word>
			<word pos="VVFIN">heiU+00DFe</word>
			<word pos="BRACKET">[</word>
			<word pos="NE">John</word>
			<word pos="BRACKET">]</word>
			<word pos="FULL">?</word>
		</sent>

		<sent>
			<sent_lang>English</sent_lang>
			<full_sent>My name is [John].</full_sent>
		</sent>

		<sent>
			<sent_lang>Irish</sent_lang>
			<full_sent>[John] is ainm dom.</full_sent>
		</sent>
</multi_lang_sent>

</subsection>
<Localised_settings>

	<Localised_setting>
		<LS_name>Integration</LS_name>
		<note>
			<note_type>PSC reference</note_type>
			<note_detail>Gaeilge--U+00C9isteacht. labhairt, lU+00E9itheoireacht,
scrU+00EDbneoireacht >tU+00E9amaU+00ED: MU+00E9 fU+00E9in</note_detail>
			<note_detail>English-- oral language, reading and writing>competence
and confidence in using language and receptiveness to language</note_detail>
			<note_detail>History--Myself and my family or local studies--
Myself</note_detail>
```

```
                        <note_detail>SPHE--Myself--Self-identity</note_detail>
                        <note_detail>SPHE--Myself and Others--Relating to others</note_detail>
                </note>
        </Localised_setting>
        <Localised_setting>
                <LS_name>Linkage</LS_name>
                <note>
                        <note_type>German section link</note_type>
                        <note_detail>Section Four- General Introductions- more advanced
1</note_detail>
                        <note_detail>Section Five- General Introductions- more advanced
2</note_detail>
                        <note_detail>Section Six- Finding out what things are
called</note_detail>
                        <note_detail>Section Seven- Family</note_detail>
                        <note_detail>Section Twelve- Pets</note_detail>
                </note>
        </Localised_setting>
        <Localised_setting>
                <LS_name>Language Awareness</LS_name>
                <note>
                        <note_type>Comp</note_type>
                        <note_detail>Examine HELLO expressions in German, English and Irish.
Note dia duit and its response(s).</note_detail>
                </note>
                <note>
                        <note_type>Theme</note_type>
                        <note_detail>Examine dialects in Irish - three ways of asking HOW ARE
YOU?-  Chaoi a bhfuil tU+00FA? Conas atU+00E1 tU+00FA? Cad U+00E9 mar atU+00E1
tU+00FA?</note_detail>
                </note>
                <note>
                        <note_type>Gram</note_type>
                        <note_detail>Word order for naming yourself in Irish differs from
German and English - [] is ainm dom, ich heiU+00DFe [], my name is [].</note_detail>
                        <note_detail>Examine contractions in English (e.g. how&amp;apos;re/how
are), Irish (e.g. TU+00E1im/TU+00E1 mU+00E9) and German (e.g. geht&amp;apos;s/geht
es)</note_detail>
                </note>
                <note>
                        <note_type>Skill</note_type>
                        <note_detail>Examine how questions are asked in each of the
languages.</note_detail>
                </note>
        </Localised_setting>
        <Localised_setting>
                <LS_name>Common errors for this section</LS_name>
                <note>
                        <note_type>Top 5</note_type>
                        <note_detail>1. Written form-Incorrect case</note_detail>
                        <note_detail>2. Difficulties with pronunciation</note_detail>
                        <note_detail>3. Adjective-noun agreement (case/gender)</note_detail>
                        <note_detail>4. Written form-Spelling error</note_detail>
                        <note_detail>5. Understanding</note_detail>
                </note>
                <note>
                        <note_type>Integ</note_type>
                        <note_detail>The -Guten Morgen, Guten Tag, Gute Nacht- combinations
provide the first introductions to Noun capitalisation in German, noun gender and/or Adjective-Noun
greement.</note_detail>
```

```
                    <note_detail>The -Ich heiU+00DFe [John]- language of this section is the
first introduction to verbs, and their inflection</note_detail>

                </note>
            </Localised_setting>
</Localised_settings>
</section>
……..
```