

**Aspects of Cohesion in Web Site Translation:
A Translator's Perspective**

Annette Schiller BA (Mod), MA

PhD

Dublin City University

Prof. Jenny Williams

School of Applied Language and Intercultural Studies

September 2008

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of PhD 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:

ID No.: 52164845

Date: 19 September 2008

Acknowledgements

First and foremost my grateful thanks are due to my supervisor, Prof. Jenny Williams for her encouragement, advice and guidance throughout this project. Thanks also to my colleagues in CTTS and DCU for their support in various ways: Mary Phelan, Dr. Minako O'Hagan, Prof. Michael Cronin, Dr. Gabrielle Carty and Dr. Claire Bohan. To Mai and Shem who are always available for discussion and debate. Last but not least, a big thank you to Willi, Isabel and Robin - without your support this project would never have been started or finished.

Finally, I would like to acknowledge the SALIS scholarship which was granted to me by the SALIS Research Committee and without which this research project would not have been possible.

Table of Contents

Declaration	ii
Acknowledgements	iii
Table of contents	iv
Abstract	viii
Abbreviations and definitions	ix
Chapter I - Introduction	1
1.0 Introduction	2
1.1 Background	2
1.1.1 Rationale	2
1.1.2 Cohesion	10
1.1.3 Cohesion and translation	17
1.1.4 Objectives	18
1.2 Methodology and theoretical framework	18
1.2.1 The corpus	21
1.2.2 The search engine	24
1.2.3 Scope of the research project	25
1.3 Structure of the dissertation	26
Chapter II – Cohesion and web sites	28
2.0 Introduction	29
2.1 Traditional models of cohesion	29
2.1.1 Halliday and Hasan	29
2.1.1.1 Reference	31
2.1.1.2 Substitution / Ellipsis	33
2.1.1.3 Conjunction	34
2.1.1.4 Lexical cohesion	35
2.1.2 Hoey and lexical cohesion	37
2.1.3 Martin	38
2.1.3.1 Conjunction	38
2.1.3.2 Participation identification	39
2.1.3.3 Lexical relations	40
2.1.4 Martin and Rose	40
2.1.3.1 Conjunction	40
2.1.3.2 Identification	41
2.1.3.3 Taxonomic relations	41
2.1.4 Halliday and Matthiessen	42
2.2 The web site	44
2.2.1 Web sites and hypertext	50
2.2.2 Analysing multimodal documents	55
2.2.2.1 Finding web content using a search engine	59
2.2.2.2 Web site navigation	60
2.2.2.3 Interaction with on-page elements	61
2.3 The relevance of traditional models of cohesion to a web site	64

2.4 Cohesion and visual composition	69
2.4.1 Framing	70
2.4.2 Information value	73
2.4.3 Saliency	75
2.5 A definition of web-specific cohesion	76
2.6 Approaches to the translation of web sites	78
2.6.1 Machine translation	78
2.6.2 Web site localization	78
2.6.3 Translation company	80
2.6.4 Freelance translator commissioned directly	81
2.7 Translation of the corpus web sites	83
2.8 Conclusion	86
Chapter III – A Model for the Analysis of Web Site Cohesion	87
3.0 Introduction	88
3.1 Background	88
3.2 Online search – an overview	90
3.2.1 The Google search engine	94
3.2.2 Search result	96
3.2.3 Search engine optimization	99
3.3 Online search – the model	101
3.3.1 Keywords	102
3.3.1.1 Keyword links	107
3.3.2 HTML tags	107
3.3.2.1 The title tag	108
3.3.2.2 The h1 Header tag	113
3.3.2.3 The description content metatag	115
3.3.2.4 The keywords metatag	118
3.4 Online search - applying the model	119
3.5 Navigation – an overview	125
3.6 Navigation – the model	127
3.6.1 Embedded navigation system	127
3.6.1.1 Embedded navigation system – global	127
3.6.1.2 Embedded navigation systems – local	131
3.6.1.3 Embedded navigation systems – contextual	134
3.6.2 Supplemental navigation system	137
3.6.2.1 Supplemental navigation system - sitemap	137
3.6.2.2 Supplemental navigation system – site index	138
3.6.2.3 Supplemental navigation system – site-specific search feature	140
3.6.3 Labels for navigation systems	142
3.6.3.1 Embedded navigation – the labels	143
3.6.3.2 Supplemental navigation – the labels	145
3.7 Navigation – applying the model	145
3.8 Page Content – an overview	147
3.8.1 Teaser text	148
3.8.2 Widgets	154
3.8.3 Information box	156

3.9 Page Content – applying the model	157
3.10 Conclusion	158
Chapter IV – The Analysis of Cohesion in Online Search	159
4.0 Introduction	160
4.1 Online Search	160
4.1.1 Analysis of keywords	161
4.1.1.1 Search using main keyword	162
4.1.1.2 Search using a keyword phrase	165
4.1.2 Issues with keywords in Online Search	166
4.1.3 Analysis of page title	173
4.1.3.1 Page titles in the SL	173
4.1.3.2 Unique page titles	174
4.1.3.3 Page titles optimized for skimming and scanning	175
4.1.4 Issues with page titles	177
4.1.5 Analysis of description content metatags	179
4.1.5.1 Google uses description content metatag	180
4.1.5.2 Google uses snippets for description	181
4.1.5.3 Google uses Web directory extracts for description	184
4.1.5.4 Description text in the source language	187
4.1.6 Issues with description content metatags	188
4.2 Conclusion	191
Chapter V – The Analysis of Cohesion in Navigation	193
5.0 Introduction	194
5.1 Navigation	194
5.1.1 Indicators of TL content	195
5.1.2 Issues with Indicators of TL content	203
5.1.3 The language of link labels and their destination content	206
5.1.3.1 SL link labels as part of TL content	206
5.1.3.2 Issues with SL link labels as part of TL content	210
5.1.3.3 Link labels in the TL with SL destination content	215
5.1.3.4 Issues with TL link labels with SL destination content	218
5.1.4 Site-specific search features	219
5.1.4.1 Search results for site-specific search features	221
5.1.5 Issues with site-specific search feature	224
5.2 Conclusion	225

Chapter VI – The Analysis of Cohesion in Page Content	227
6.0 Introduction	228
6.1 Page Content	228
6.1.1 Teaser text	228
6.1.1.1 Teaser text headings are in the SL, are ambiguous or obscure	230
6.1.1.2 Issues with teaser text headings in the SL, ambiguous or obscure	231
6.1.1.3 Teaser text headings do not correspond to their destination content	234
6.1.1.4 Issues with teaser text headings not corresponding to destination content	237
6.1.1.5 Direct address and superfluous text in the teaser text	238
6.1.1.6 Issues with direct address and superfluous text in teaser texts	242
6.1.2 Widgets	245
6.1.2.1 Widgets in the SL, ambiguous or imprecise	246
6.1.2.2 Issues with widgets in the SL, ambiguous or imprecise	253
6.2 Conclusion	256
Conclusion	254
7.0 Introduction	255
7.1 Aims and achievements of study	255
7.2 The role of the web site translator/translator training	258
7.3 Addressing issues of cohesion in the corpus web sites	262
7.3.1 Cohesion in Online Search	262
7.3.2 Cohesion in Navigation	263
7.3.3 Cohesion in Page Content	264
7.4 Suggestions for further research	265
References	266
Appendix A	1-3
Web site statistics - examples	
Appendix B	1-4
Checklist – addressing of cohesion in the translated web site	

Abstract

Annette Schiller: Aspects of Cohesion in Web Site Translation: A Translator's Perspective

This dissertation investigates the nature of cohesion on a web site and the implications for the web site translator. It approaches the subject from the perspective of the freelance translator.

Cohesion traditionally refers to the network of grammatical and lexical items in a text that combine to link different parts of a text and give structure to the text. It is one of the most challenging aspects of translation as each language has its own unique manner in which it employs cohesive devices in the creation of a cohesive text.

A web site can be described as a multi-modal and multi-linear instrument of communication where language, image and audio content combine to create cohesion. Hypertext, the defining feature of the World Wide Web has led to a change in the way in which content is accessed, in reading patterns and in the general ordering of information on a web page and web site. Three key areas are identified as being fundamental to web site cohesion, Online Search, Navigation and Page Content, which reflect changing user interaction with the web site as text. Information is now available to the user much in the way that people think, that is, by association rather than in linear sequence. As such, the web site presents a new challenge to the translator in terms of identifying and addressing items that are capable of functioning cohesively on a web site.

Traditional models of cohesion are examined and found to be inadequate for the analysis of web site cohesion. A definition of web-specific cohesion is proposed and a model is created for the analysis of aspects of cohesion that are relevant to the translated web site. The model is applied to the English-language content of a corpus of German source-language web sites. The findings are discussed, implications examined, the changing role of the freelance translator in web site translation outlined and topics for further research suggested.

Abbreviations and definitions

Body text	Continuous text on a web page
Cached view	Snapshot taken of a web page by Google
Deep-linking	In online search, where the user is directed to a page that matches the specific information required and the page is not the homepage.
Description content metatag	Text that describes the content of a given page, contains one or more of the most important keywords and is part of the HTML
Description text	In this context, the text provided in a search result by Google to describe the content of the web page to which that search result links
Domain name	A name that identifies a web site, consisting of a specific part (on the left) and a more general part (on the right, e.g. muenchen.de
DTS	Descriptive Translation Studies
HTML	HyperText MarkupLanguage - The structure, layout and look of a web page, including the colour, fonts and font sizes are controlled using HTML, which also supplements the text content with interactive elements and embedded images
Keyword	The most important term or terms used to describe the content of a web page or site which are also likely to be used by the user in online search
Keyword link	A keyword link refers to a link label that consists in whole or part of a keyword or keywords

Keyword phrase	Two or more keywords that commonly co-occur in a given context
Keywords metatag	Contains a list of the most important keyword defined for a given page, is contained in the HTML title tag, is not displayed on the web page
Link label	The label (textual or iconic) given to a link and representing the destination content of that link
Page title	Contains the most important keyword(s) that describe the content of a given page, is contained in the HTML title tag
SC	Source culture
SE	Search engine, e.g. Google
SEO	Search Engine Optimization - the process of developing a website's structure and content to enable it to achieve the best possible rankings within the results being displayed on a search engine after a search is conducted for a particular term
SERP	Search Engine Results Page(s) - List of results generated by a search engine in reply to a request for information by a Web user
SL	Source Language
ST	Source Text
TC	Target culture
Teaser Text	Short blurb that is used to indicate the presence of additional content on a web site. It is partly or wholly hypertextual.
Title tag	HTML tag that contains the page title for each page
TL	Target Language

TS	Translation Studies
TT	Target Text
URL	Uniform Resource Locator – web page address
Web	World Wide Web / WWW
Widget	Created from the words ‘window gadget’, it refers to a user interface to a program that presents information using windows, menus, buttons, text fields. Flight and hotel booking forms are commonly-used widgets.

Chapter I

Introduction

Introduction

The vision I have for the Web
is about anything being potentially
connected with anything

Berners-Lee (1999 p1)

1.0 Research question

This study seeks to investigate cohesion in the translated content of multilingual web sites. Traditional views of ‘cohesion’ are used in the context of the print medium and refer to the unity of a text that is brought about by the connectedness to each other of individual elements of that text. In this thesis it is hypothesized that the advent of hypertext has led to the need to revisit and redefine cohesion in a web site, which in turn has implications for cohesion of the translated content of web sites. It is further hypothesized that the failure to adequately address cohesion in the translated web site has implications that are potentially more far-reaching than would be the case with a text in the print medium.

1.1 Background

1.1.1 Rationale

In the early days of the World Wide Web, it was thought that the language of the Web would forever be English, the language of the first web sites. While English is the language of the greatest percentage of Web users at 30.4% of the total online population (Internet World Stats 2008), statistics show a considerable increase over the past number of years in speakers of languages other than English accessing the

Web. Almost one billion users of the Web today are non-English speakers. The number of Chinese speakers accessing the Web, for example, rose by over 600% between 2000 and 2008, and now stands at 233 million. This is compared with 427 million English speakers and 122 million Spanish speakers (Internet World Stats 2008).

The benefits to an exporting organisation of offering foreign-language content on its web site are well documented (LISA/AIIM 2001; Ramsey et al 2002; O'Hagan and Ashworth 2002; CSO 2006). One of the basic principles of marketing, that of addressing the customer in his/her own language, is summed up by O'Hagan and Ashworth:

In order for a business to take advantage of the emerging worldwide marketplace on the Internet, it clearly needs to speak the customer's language and take into account some of the cultural factors relevant to that particular market (2002 p 12).

Many companies and other organizations are attempting to accommodate this demand for information in the language of the client user by providing content in the language or languages of their target export markets, therefore enabling, customers to 'read text and navigate in their own language' (ibid.).

As the Web has become increasingly part of our daily lives, user behaviour on web sites and the general attitude to and acceptance of the Web have changed. Nielsen and Loranger (2006 p xx) suggest that in the early years of the Web users were 'grateful' if they found the information they were looking for. These days, however, expectations are greater than ever before. Users assume that the information they wish to access is available on the Web and they also assume that they will be able to

find and access that information through a well-functioning and user-friendly web site (Nielsen and Loranger 2006 p xx.). Likewise, the user wishing to access the foreign language content of a web site will expect this content to be just as user-friendly as the source language (SL) content s/he may have used to date (O'Hagan and Ashworth 2002 p ix). As is the case with the print medium, the structure of a text and the connectedness to each other of the surface elements of a text, the cohesion of a text so to speak, is a significant factor in enabling the user to access the information required and to move between items of information on a web site. It is the aim of this dissertation to examine the nature of web site cohesion and specifically to explore the issues that arise for cohesion in the translated web site, and specifically in the 'information web site'.

My interest in the subject of cohesion in the translated web site from the point of view of the freelance translator stems on the one hand from my work as a professional translator and on the other from my general involvement in the profession as Chairperson of the Irish Translators' and Interpreters' Association (ITIA). From the point of view of the translator today, the Internet is significant on a number of levels. It is a valuable resource for terminology and subject area research on both mono and multilingual web sites, for maintaining contact through e-groups or discussion forums with other translators working in the same language or subject area and for receiving and returning assignments by email, regardless of the geographical location or time zone of either translator or client. As a freelance translator with clients in Ireland, Germany and Switzerland, my own personal involvement in web sites is both as a frequent user of multilingual sites for research

purposes and as a translator of web sites. My experience as a freelance translator of web sites is varied. Generally, the source text (ST) is provided in a Word file, the target text (TT) is returned in Word and is uploaded by an IT person to the web. Sometimes the insistence that the (TT) be reviewed *in situ* before it goes live on a web site is heeded, at other times the translator has no influence over the text once it has been returned to the client. At times the translator can use the source web site as a guide for the purpose of layout, colour scheme, non-verbal elements and overall site structure. At other times, the source web site may not yet be available. It is also not uncommon for the translator to be unaware that a TT will be used on a web site. This can happen, for example where a text that has been translated specifically for a company printed brochure is used for the company web site. Baldry and Thibault (2006 p156) write that ‘many existing forms of hypertext are banal, commercialised recyclings of already existing textual forms, which are simply uploaded to a website’.

As Chairperson of the ITIA, the researcher is in regular contact with others in the profession both in Ireland and abroad, through sister associations in Europe, the US and New Zealand. These contacts include translators, the vast majority of whom are freelance, translator trainers, translator companies and translation clients. In personal communications with practitioners and translation companies it has emerged that translator experiences are similar across Europe. Firstly, the translation of web sites is more often than not carried out by freelance translators (Boardman, DCU-LS, Eymard-Duvernay, Heil, Minogue all 2008). This is supported by the results of a recent survey (results forthcoming) carried out by the ITIA among its members, 90%

of whom are freelance translators. The survey contains a question on the involvement of the members in web site translation. Only 22% of the respondents never translate web sites as part of their translation work while 39.5% translate web sites 'frequently'. Communications with those in the profession also suggest that web site translators work from different file formats (e.g. XML, HTML), but in the majority of cases from a Word file. The freelance translator works either directly for a client or for a translation company. The involvement and influence of the translator, therefore, varies from assignment to assignment and from client to client, and will depend among other things on the experience of the translator, the attitude of the client and his/her understanding of the importance of target language (TL) content as part of the overall online presence of the company.

Given that over 90% or about 19 million of all companies in the EU are in the Small and Medium-sized Enterprise (SME) sector, it is reasonable to assume that the web site translation carried out by freelance translators in Europe is, in the main, for the SME sector. The importance of the Web for this sector has long been identified:

small and medium-sized enterprises (SMEs) have been identified as key users of Internet commerce, ... there is increased speculation that SMEs will be able to capture the global markets, sell to international customers and compete favourably with large Corporations' (Ramsey et al 2002 p119).

However, due to the precarious financial situation of many exporting SMEs (Scharf et al. 2004 p103), the provision of foreign-language content is not always considered to be a priority for such companies, particularly where the cost of setting up the original, SL web site may have been quite considerable. The difficulty on the part of web owners of appreciating the value and importance, firstly of a multilingual

presence and secondly, of well-translated and cohesive TL content that is on a par with the source web site, became evident during my work as a consultant reviewer of the export communication needs of SMEs in Ireland (pilot scheme carried out by Chambers Ireland under the auspices of the European Leonardo da Vinci/Protocol II programme). The web site text for translation is treated, therefore, much as if it were in the print medium, and indeed the expectation is that it can simply be translated as a text in isolation from the site and uploaded to the web site. According to O'Hagan and Ashworth, the most common notion among translation clients is that

'Translation is a stand-alone element that can be added on, and is fundamentally separate from the overall inter-lingual communication process' (2002 p3). Such an approach results in the textual elements of a web page, be it links, a navigation bar or continuous text being translated, page for page without due consideration of their overall contextual situation in the network of the web site and with little or no awareness of the consequences on reading patterns in a hypertextual environment.

The TL content is not considered to be valuable web site content in its own right and with its own requirements in terms of cohesion.

Personal communications with those in the profession (Heil, Boardman all 2008) also indicate that web sites are frequently translated by non-translators. The IT person dealing with the source web site is sometimes asked to provide the foreign-language content. In addition, once the initial TT has been provided to the web owner it is not uncommon for the administrative personnel at the client company to be expected to provide any updates that may be required in the TL, (Esser, Hartlaub, Hoppe, Rappenglück all 2008). Consequently, the TT may not receive the same

attention to detail that the ST may have received, which includes for example, recognising the nature of cohesion and cohesive devices on the source web site and providing for a TL web site that can also function cohesively.

One of the issues resulting from a lack of cohesion on a web site concerns the visibility of the TL content, not only where the user is attempting to access such content through a search engine (SE) such as Google or Yahoo! but also where the user is already on a multilingual web site but the presence of the TL content is not signalled or not signalled adequately. Content that is relevant and potentially of high value to the user may not be found.

A second issue that is regularly encountered on multilingual web sites concerns the situation where the user typically follows one link after the other in the TL, moving through a web site, closer to the content that is the objective of the web site visit, only to arrive at the destination page to find that the page is in the source language (SL) due to the fact that that particular piece of information or that page has not been translated into the TL or an incorrect address has been placed in that final link. In this case, there can be no cohesive link between one item and the next. This failure to facilitate a cohesive web site could result in the user losing all sense of context and orientation due to the SL environment. Multilingual web sites that have issues such as those described above display little sense of unity or cohesion and offer a low level of user satisfaction. The implications are serious. The user wishing to access the TL content on such a site, and who is the very reason the web site owner decides to provide TL content in the first place, will often not persevere if such content is difficult to access. The credibility of a web site will suffer as a result. The

user may leave the web site to search for and access another site with similar content. Speed is of the essence in an online environment and a cohesive web site will enable the user to make quick decisions and move swiftly back and forth through a site. If this is not the case, the overall traffic to the TL content of a multilingual site may not be as considerable as could be expected. This in turn could convince the web site owner that the TL content is not as valuable or popular as initially expected and should not, therefore, be the subject of undue investment or attention in the future.

This dissertation addresses the nature of cohesion on a web site, the resulting linguistic challenges facing the freelance translator of web sites and the decisions that are required of him/her in terms of providing a cohesive text. It is also concerned with the changing role of the translator in web site translation. Addressing the issue of cohesion in the translated web site will, for example, involve the translator in an approach that encompasses what Rosenfeld and Morville refer to as 'Information Ecologies' (2002 p24), including the three items 'user', 'context' and 'content'. The translator will not simply deal with a text in isolation that has been extracted from the source web site. S/he will be familiar with and understand the language of the potential TL 'user' of the site, in the particular 'environment' in which the company or business functions and will be familiar with how the 'documents, applications, services and metadata that people need to use or find (a) site' (ibid. p25), therefore providing translated web content that can be considered to be an integral part of the overall site and not merely an appendage. The role of the

freelance translator in the translation of web sites and the issue of translator training are discussed in more detail in the concluding chapter of this dissertation.

1.1.2 Cohesion

Halliday and Hasan were the first to define ‘cohesion’. Cohesion is traditionally used in the context of verbal text and is expressed ‘partly through the grammar and partly through the vocabulary’ of a text (Halliday and Hasan 1976 p5-6). It

does not concern what a text means; it concerns how the text is constructed as a semantic edifice (ibid. p26).

Cohesion occurs where

the interpretation of some element in the discourse is dependent on that of another. The one presupposes the other, in the sense that it cannot be effectively decoded except by recourse to it (ibid. p4).

It refers to the network of lexical and grammatical relations in a text that connect different parts of that text, creating a unified whole. Consistency of style and tense and the use of punctuation can also contribute to the cohesion of a text (Baker 1992 p193). Cohesion gives structure to a text and enhances the readability and coherence of a text. Cohesion in Hallidayan linguistics applies to linear text, where ‘the sentences of a text can only follow one after the other’ (ibid. p227). Such text for the purposes of the current study on cohesion in web site translation is referred to as ‘conventional text’, meaning that it is ‘printed/written on a page (a sense of permanence), linear (read from beginning to end) or subject to some kind of external “editing” (quality control)’ (Bowker 2005 p10).

Cohesion is undoubtedly one of the most challenging areas of translation, given that the cohesive devices used, the frequency with which they are used, and the cohesive

patterns created in a text vary from one language to another. The use of cohesive devices can also vary within the same language depending on genre and text type (Smith & Frawley 1983 p371). In particular, lexical cohesive patterns can rarely if ever be replicated in the TT (Baker 1992 p206). Consequently, the translator must find ways of replacing these patterns in a way that will enable the TT to retain the same level of cohesion as the ST. A translated text that contains only source language cohesive patterns will not be capable of functioning cohesively in the target language and culture.

A web site can be described as a multi-modal instrument of communication, multimodality being

the use of several semiotic modes in the design of a semiotic product or event together with the particular way in which these modes are combined – they may for instance reinforce each other ... , fulfil complementary roles ... or be hierarchically ordered (Kress and van Leeuwen 2001 p20).

Meaning on a web site is created through language, image and audio content and the unity or cohesion of a web site, therefore, comes about through a combination of these modes. The web site as ‘text’ provides an entirely new challenge to the translator in terms of cohesion as it represents a departure from what we know as conventional text. Baldry and Thibault (2006 p104) refer to the web page as a ‘visual-spatial unit on a computer screen’. The web page can also be classed as ‘hybrid’ given that it contains not only features similar to those of the printed page (language, resources of depiction, spatial juxtaposition of objects) but is in addition hypertextual in nature and offers potential for action, where ‘the user can *act* on the page and obtain *responses* to his or her actions’ (ibid.).

Despite the multimodal nature of the web site as text, this research will focus on those elements of the web site, with which the translator is primarily concerned, namely verbal language and the linguistic decisions required to create a TT, and specifically cohesion in the TT. This approach is supported by the belief that language is central to the Web (Crystal 2001, Ravichandran et al 2004, Nielsen and Loranger 2006).

Text is present on a web site in many forms, for example as body text, hyperlinks and HTML text. Unlike cohesion in a conventional text, the nature of cohesion of such diverse types of text is influenced in particular by two factors. The first is the medium with which web site content is accessed. Text on a web site is generally read on a screen, in the majority of cases on a PC or laptop and increasingly on different types of handheld portable devices such as the BlackBerry or the iPhone. Text on a screen cannot be physically read as quickly as print text which has implications for the way in which the user reads text on the screen (Nielsen and Loranger 2006 p234) and consequently for the way in which such text is laid out and ordered. For example, large chunks of text are difficult to read on a screen, are therefore less likely to be read in full from start to finish, and are frequently divided into smaller chunks, despite the fact that the chunking may go against standard conventions in the language in the print medium.

The second and most important factor that influences cohesion in web text is hypertext, which is defined as

generally, any text that contains links to other documents - words or phrases in the document that can be chosen by a reader and which cause another document to be retrieved and displayed (Jaha Design 2007).

Hypertext provides the basis for the multi-linearity of a web site and has changed the way in which people find, access and read content. Thanks to hypertext the reader is no longer 'constrained to read in any particular order but (can) follow links and delve into the original document from a short quotation' (Berners-Lee 1999 pp5-6).

Reading patterns are no longer text-driven but user-driven.

With hypertext, the actual text gets constructed as the reader selects which relationship(s) to explore as she chooses hyperlinks and reads the chunks of text in the nodes she selects. As a result, the meaning the reader makes of the hypertext depends in part on which nodes she visits, and in what order (Cripps 2007).

Baldry and Thibault (2006 p116) refer to the pathway selected by the user as a 'trajectory':

A meaning-making trajectory in this sense refers to the progressive integration over time of the semiotic resources that are encountered as the website user progresses from one linked object, one text, one web page and one website to another.

A trajectory can last a matter of seconds or considerably longer and can also be terminated and re-entered at a later point in time (ibid.). The process of searching for specific web content and accessing it directly through online search is not an inherent part of a reader's engagement with text in the print medium. In addition, the success or otherwise of an online search can be influenced to a considerable extent by the web site text itself, where a text can be 'prepared' in order to increase its visibility for SEs.

Three key areas of user interaction with web sites emerge that set it apart from conventional text. These three areas involve firstly, the use of a SE by the user for

seeking and accessing online content, an activity that does not apply to print media. The user can enter the site on any page of a web site, even the most obscure, which means that each page of the site must be capable of providing a sense of context and orientation, through cohesion. Secondly, the creation or even re-creation by the user of reading paths on a web site or between web sites, where there is 'no single, determinate starting point or sequential organisation' and where 'the locus of control shifts away from a stable genre schema to the computer user as the maker and improviser of solutions' (Baldry and Thibault 2006 p117), is a departure from the linearity of conventional text. The third area concerns the interaction of the user with different on-page elements such as 'widgets', where the user provides information in order to receive information.

Overall, much is expected of the web site user in a short space of time in terms of finding and accessing the required web content, understanding how to navigate a site once s/he has accessed the site and also understanding how to handle and use the various items of content as presented on the relevant pages of a site.

As is the case with any text, a cohesive text will enable the user to move swiftly from one item to another, and in the case of a web site, from SE to web site, and will contribute in no small way to a good user experience.

The fact of hypertext and the reading pathways resulting from it would suggest that traditional, linear-based, approaches to cohesion are inadequate, in their entirety, to describe and explain the nature of web site cohesion, where cohesion is seen to express 'the continuity that exists between one part of the text and another' (Halliday and Hasan 1976 p299).

This dissertation will demonstrate that the success of textual cohesion on the translated web site is dependent not only on textual elements that are capable of functioning cohesively, but also to a great extent on the interrelationality of modes, that is to say on visual composition to support cohesion. The principles of visual communication apply not only to images but also to the layout of verbal text in a visual composition. It will be seen that while language is naturally governed by the 'logic of time and by the logic of the sequence of its elements in time, in temporally governed arrangements' (Kress 2003 pp1-2), language on a web site is also governed by 'the logic of space, and by the logic of simultaneity of its visual/depicted elements in spatially organised arrangements' (ibid. p2).

The subject of textual cohesion is well documented in the context of print media. However, given the importance of the subject of web site cohesion, a review of the relevant literature has surprisingly failed to produce any literature either on this subject or specifically on the subject of cohesion and the translated web site. This study seeks, therefore, to bridge this gap by proposing a definition of web site cohesion, with the focus on those aspects that are relevant to the web site translator. In addition, this research will create a new model for the analysis of cohesion in the translated web site.

Web site cohesion consists of a network of lexical items, referred to as specifically-defined keywords that link various parts of a web page and web site in a multi-linear fashion, providing context and a sense of orientation on all pages of the web site. A keyword is a word or phrase used when searching for Web content using a SE and

which the webmaster uses when describing or explaining the content of a web page or web site (Linking101 2008). A 'keyword phrase' consists of two or more keywords used when searching for content with a SE. The cohesive potential of keywords lies in the 'repetition' of these keywords and keyword phrases. Above all, the keywords and keyword phrases reflect user terminology, which refers to the terms used by the user when searching for and accessing content on a web site. The keyword-carrying text includes HTML text (page title tag and description content metatag), hyperlinks (navigation bar, links) and page content (headings, teaser text, body text). As mentioned previously, a defining feature of web site cohesion is the fact that the verbal elements that are capable of functioning cohesively are dependent for their cohesive effect on the logic of visual communication. Similar to the print media, the challenge for the web site translator is to recognise how cohesion is created in the ST and to find suitable solutions for ensuring an appropriate level of cohesion in the TT.

1.1.3 Cohesion and translation

Cohesion is one of the greatest challenges in translation given that languages differ greatly in the cohesive devices they use, the frequency with which they are used and also their genre-specific use within a language. Lexical chains that occur naturally in the ST, creating cohesion and providing texture, can rarely if ever be replicated in the TT (Baker 2002 p 206). The difficulty of recognizing and successfully dealing with elements of lexical cohesion and particularly collocations is a constant

challenge to the translator and one which demands various strategies if the same or similar level of lexical cohesion is to be achieved in the TT as is present in the ST.

The effect of lexical, especially collocational, cohesion on a text is subtle and difficult to estimate. With grammatical cohesion the effect is relatively clear: ...In lexical cohesion, however, it is not a case of there being particular lexical items which always have a cohesive function. EVERY lexical item MAY enter into a cohesive relation, but by itself it carries no indication whether it is functioning cohesively or not. That can be established only by reference to the text (Halliday and Hasan 1976 p288).

One underlying reason for this is the issue of instantial meaning of a text whereby 'the text provides the context for the creation and interpretation of lexical relations, just as the lexical relations help to create the texture of the text' (Hoey 1991 p 8).

Snell-Hornby points out that the translator is not concerned with

isolating phenomena or items to study them in depth, but with tracing a web of relationships, the importance of individual items being determined by their relevance and function in the text (1988 p69).

The translation of idioms and other set phrases is one example of the difficulty facing the translator in terms of the attempt to re-produce a similar or equivalent cohesive lexical network. The ST idiom may at times have to be abandoned in the TT as there is no identical idiom available in the TL which can be reproduced with the same effect as produced with the ST.

Frequently, a ST collocation may not be recognised as such. This can lead to a failure to understand and transfer the intended meaning, resulting in a literal translation of the individual lexical items that will more than likely deviate from the original meaning. The cohesion that should have been recreated through the lexis of

the TT will now not come about. Baker (1992 p52) writes that decisions on collocation in translation should be influenced by register or genre, as collocations that sit well in one particular register or genre do not necessarily do so in another.

1.1.4 Objectives

This study aims to examine the nature of cohesion on a web site and to identify those devices that are pertinent to cohesion in the translated web site. As no model exists for analysing cohesion on the translated web site, a model will be created that builds on traditional models of cohesion and also draws on approaches to the analysis of visual composition, in addition to well-established guidelines on Search Engine Optimization (SEO), Information Architecture and Web Usability.

1.2 Methodology and theoretical framework

Traditional models of cohesion are examined, in particular Halliday and Hasan (1976), which is arguably the best known and most influential model in Translation Studies (House 1981, Hatim and Mason 1990, Baker 1992), in addition to the Hoey model of lexical cohesion (1991), Martin (1992), Martin and Rose (2002) and Halliday and Matthiessen (2004).

The general features of a web site are outlined and, from the translator point of view, textual elements in this hypertextual environment in particular. Against the backdrop of cohesion, the consequences of hypertext in comparison with conventional text are outlined. Three key aspects of a web site are identified, which for the purposes of this project are referred to as *Online Search*, *Navigation* and *Page Content*. They

reflect the most significant areas of user interaction with a web site that differ from user engagement with the print media and that require the re-examination of the meaning of cohesion on a web site.

The relevance of traditional models for the analysis of cohesion on a web site is examined. It is suggested that, while such models of cohesion are relevant, to an extent, for the analysis of cohesion of continuous text (also referred to as body text) on a web site, they are not sufficient to account for cohesion in the case of all instances of text on a web site.

Given the multimodal nature of the web site, approaches to the analysis of cohesion in the visual composition are considered, in particular the Grammar of Visual Design (Kress & van Leeuwen 1996), in addition to Baldry and Thibault (2006) and Bateman (2008). A definition of cohesion on a web site is proposed as a basis for exploring the implications for the translator. As no model exists to test the cohesion of a translated web site, a model is created which takes into consideration the three areas of user activity as previously mentioned. The model builds on traditional models of cohesion. It goes beyond the confines of Translation Studies (TS) to draw on models for the analysis of visual composition in addition to aspects of well-established approaches to Search Engine Optimization, Information Architecture and Usability, which are of direct concern to cohesion in the translated web site.

The model is tested on the English TL content of the corpus web sites. The most significant findings are outlined and implications for the translator are discussed.

Recommendations are made for addressing the issues of textual cohesion in web site

translation that emerge in the analysis of the corpus web sites. The changing role of the freelance translator in web site translation is discussed along with the general issue of training translators in web site translation. Finally, suggestions are made for further research in the area of web site cohesion and the translated web site.

In multilingual web site communication, the content of the TT frequently mirrors that of ST. However, at times the ST is adapted in some way before translation, for example is abbreviated, or an entirely different text on the same subject is provided for translation for the target audience. Nonetheless, all such texts presented in the TL content of the corpus web sites are considered to be translations, concurring with Toury that a 'translation' is 'any target-language utterance which is presented or regarded as such within the target culture, on whatever grounds' (1985 p20). The analysis of the corpus is non-evaluative, within the framework of Descriptive Translation Studies (DTS). As the focus of this study is essentially content-analytical and, therefore, qualitative, a corpus of ten web sites is deemed to be sufficient to observe and explain certain phenomena. The translated text is observed as a text functioning in the target or recipient culture, which 'serves as the initiator of the decision to translate' (ibid. p19). It was neither necessary nor desirable within the confines of this research project to analyse all sections and pages of a web site in order to establish the nature of cohesion and the success or otherwise of cohesive devices. Therefore, it was decided to concentrate not only on individual pages but also on subject matter, for example topics suggested by the main keywords used in

searches connected with the ten websites. This included, for example indicators of foreign-language content on a web site. The section of the model that addresses global access to foreign-language content examines firstly the homepages of the German source web sites and secondly random pages of the ST, in order to reflect the fact of random entry to a site on any one page. With regard to the section of the model on SE optimization, the search results that are returned by Google in addition to their destination pages are examined.

As previously mentioned, this project does not address translation quality or seek to evaluate or prescribe but seeks to describe and explain. However, the implications of instances where the issue of cohesion has not been adequately addressed in the TT are discussed.

1.2.1 The Corpus

In order to ensure a suitable and comparable corpus of web sites for this research project, a number of criteria were established for the selection of the corpus. The web sites had to have source and target language content, in order to fulfil the requirement for a ‘translation’. While a multilingual corpus would, in the strict sense, involve upwards of three languages and a ‘bilingual’ corpus two languages (McEnery and Xiao 2008 p19), the use of the term ‘multilingual’ in this project is used in the broad sense, that is to refer to a corpus that, in addition to source content has target language content in one or more languages. Each of the sites has content in at least one foreign language in addition to content in the source language.

The source language was to be German and the web sites could be from Germany,

Austria or Switzerland. However, it was found that there was a greater choice of web sites available from Germany given the size, population and number of companies and organisations in Germany compared with Austria and Switzerland. The web sites were to be from the same sector offering similar information on products or services, therefore allowing for comparable results. While this would normally narrow the target audience somewhat, the subject matter of the web sites chosen – all aspects of information on the city for the tourist or person living in or relocating to that city – meant that the target audience is in fact very broad. This has implications in translation for issues such as register and terminology choice. Finally, given the ephemeral nature of information on the Web, a sector had to be chosen in which the web sites were likely to be active and accessible for the medium to long term, or at the very least for the duration of this project.

The type of web site and sector was selected in line with a) the most popular type of web site activity recorded by the Central Statistics Office (CSO) in Ireland between June 2004 and June 2005 and b) the type of site with regard to that activity that was most likely to be visited by the Web user. According to the statistics issued by the CSO, the most popular Internet activity among users in Ireland related to ‘information search and online services, which refers to finding information about goods and services, travel and accommodation’ (2006 p28). In addition, the most popular type of goods and services ordered online were ‘Travel and holiday accommodation’. Short city breaks are the fastest growing tourist product for Irish people today with European cities being among the most popular destinations. It is,

therefore, reasonable to assume that the average traveller who books a holiday online will also carry out some research online about the holiday destination, preferably in his/her own language. It was decided, therefore, to use a corpora of sites that would provide information for the average short-break tourist from Ireland on the destination cities in Germany. The official web sites of ten cities were chosen: Munich, Frankfurt, Berlin, Hamburg, Stralsund, Regensburg, Garmisch-Partenkirchen, Leipzig, Münster and Cologne. 'Official' refers to the fact that the web sites are set up and administered on behalf of the city, for example NetCologne is the company that runs the site for the city of Cologne. The official sites were chosen because it is believed that they would carry more weight and therefore have a greater level of credibility than a purely commercial site. Such a level of credibility brings with it a certain expectation in terms of providing a quality, user-friendly and cohesive web site. The corpus web sites offer content on broadly the same general subjects, that is to say content that is of interest to tourists and also people resident in the city or people relocating there. The corpus web sites represent a broad geographical spread within Germany. Large cities such as Berlin and Hamburg are represented alongside smaller cities such as Regensburg, Garmisch-Partenkirchen, which are also important from a tourism point of view. The target audience for the translated content is a native English speaker and a speaker of English as L2. Given the fleeting nature of content on a web site as mentioned, it may happen that some of the examples presented in this study, while online and valid at the time of data collection and evaluation, are no longer accessible or have in the meantime

been modified in some way. This is the case, for example, with the analysis of 'Online Search' in Chapter IV. The information on the keywords 'munich' and 'restaurant' is no longer valid as this section of the web site has been revised to include the keyword 'restaurant' on the main English restaurant page.

1.2.2 Search engine

There are numerous SEs in use worldwide today, however, the decision was made at an early stage of this project to use one SE for the purposes of research into online search in the context of cohesion and the translated web site. The use of a single SE ensures a high level of consistency in terms of search results, given that SEs have different methods of extracting information from web sites, of indexing this information, evaluating web pages and of providing information in a search result. In many cases the exact means of providing information in a search result is not known outside the individual company. For example, the exact composition of the search algorithm used by Google when deciding on the page ranking of a web page is only known within the company.

The SE Google was chosen for the study for a number of reasons. Firstly, Google is the most popular SE among the many SEs available for English-language Web users worldwide. In October 2007, it achieved a US market share of 55.5%, while the next most popular SE, Yahoo!, had a market share of 18.8% (Nielsen//NetRatings 2007). In the UK, Google was the most popular SE in August 2006 with 68% of the market, compared with Yahoo! in second place with 9% (Nielsen//NetRatings UK 2006c).

While no figures for Ireland are available, it seems reasonable to assume that user figures in Ireland for search using the Google search engine are broadly in line with the figures from the US and the UK.

A second reason for using the Google search engine is that the methodology used by Google in choosing and displaying web pages for its SERPs while never fully transparent is somewhat more transparent than that used by other search engines. It, therefore, lends itself better to the analysis of search engine requirements, allowing both the web writer and in the case of this study, the translator, to address such optimization techniques as may be appropriate. Finally, while each search engine has its own individual approach to web page ranking and the creation of search results, many of the issues addressed in the context of the Google can be applied to the translation of web sites even where that site is not optimized with Google in mind.

1.2.3 Scope of the research project

This study does not aim to address the subject of web site translation or that of web site cohesion *per se*. It does not provide an exhaustive list of cohesive devices relevant to the web site but examines those cohesive devices that are of direct concern to the translator in terms of non-continuous text on a web site. While it is understood that the translator may be required to address non-verbal (e.g. image) elements on a web site and how they combine to create meaning, this research addresses only the issues that arise in connection with cohesion of the verbal text. In addition, this research deals with translation as a product and with the implications for the freelance translator. It does not purport to be a guideline on how

to translate text on a web site or how to deal with different programmes, file formats or platforms that may be involved in the creation of the web site TT. Rather it concentrates on the linguistic issues that arise in terms of cohesion and the linguistic decisions that are made by the translator in providing a cohesive TT. While this research is carried out from the perspective of the freelance translator, the recommendations discussed in the final chapter of this dissertation are applicable to all translators of web sites, regardless of whether they are part of a large localization project or are working alone with or without such translation aids as Translation Memory (TM) or TagEditor.

1.3 Structure of the dissertation

Chapter II of the dissertation outlines traditional approaches to cohesion, examines the features of a web site against the backdrop of cohesion, particularly those that are relevant to the translator, and considers the extent to which traditional approaches to cohesion are relevant for cohesion on a web site. The relevance of aspects of visual communication to the cohesion of web sites is considered and a definition of web site cohesion is proposed. Different approaches to the translation of web sites are discussed. Chapter III presents the model for the analysis of cohesion on the translated web site. Chapters IV, V and VI present the findings of the application of the model to the corpus web sites. The implications for the translator are discussed. The final Chapter in this dissertation presents conclusions, discusses the role of the freelance translator in web site translation and issues for the training of translators in web site translation. Recommendations are made for

addressing issues of cohesion that arise in the analysis of the corpus web sites and a supporting checklist is provided in Appendix B. Suggestions for further research in the area of web site cohesion are made.

Chapter II

Cohesion and Web Sites

Chapter II

Cohesion and Web sites

2.0 Introduction

Chapter II begins by outlining traditional approaches to cohesion, including that of Halliday and Hasan (1976), Hoey's model of lexical cohesion (1991), Martin (1992), Martin and Rose (2002) and Halliday and Matthiessen (2004). In addition to general observations about web sites, the nature of text on a web site is examined. The consequences of hypertext are outlined, in particular three key areas of user interaction with a web site which are referred to as *Online Search, Navigation* and *Page Content*. The relevance to web site text of traditional approaches to cohesion is considered. The relevance to web site cohesion of aspects of visual composition in terms of Kress and van Leeuwen (1996) are considered as are approaches to the analysis of multimodal texts of Baldry and Thibault (2006) and Bateman (2008). A definition of web site cohesion is proposed. The main approaches to web site translation are discussed and the individual approaches taken by the corpus web sites for the translation of their web content are outlined.

2.1 Traditional models of cohesion

2.1.1 Halliday and Hasan

In Hallidayan linguistics, cohesion is one of a number of elements that comprise the functional components that realize communication. These components or

metafunctions are referred to as the *ideational*, the *interpersonal* and the *textual*. The *ideational* metafunction expresses the content of the text. The *interpersonal* metafunction articulates the social and interpersonal relationship of the participants to each other in the given communicative situation and also the attitudes of these participants to each other. The *textual* metafunction is the ‘text-forming component’ in the linguistic system and has two components: *information structure* and *cohesion* (Halliday and Hasan 1976 p27). Information structure refers to the ‘ordering of the text, independently of its information’ on the basis of the distinction into ‘given and new’ (ibid.). The second part of the textual metafunction, ‘cohesion’ is described as

the means whereby elements that are structurally unrelated to one another are linked together, through the dependence of one on the other for its interpretation (ibid.).

Traditionally, the term cohesion is used in the context of written or spoken language to express the ‘continuity of text that exists between one part of the text and another’ (p299). It is the ‘set of possibilities that exist in the language for making text hang together; the potential that the speaker or writer has at his disposal’ (ibid. 18).

Cohesion is not concerned with meaning in a text but with the construction of the text as a ‘semantic edifice’ (ibid. p26).

The concept of cohesion is a semantic one; it refers to relations of meaning that exist within the text, and that define it as a text (ibid. p5).

‘Text’ is defined as a ‘unit of language in use’ (ibid. p1), which can be spoken or written, prose or poetry and of any length. The relations set up by cohesive devices

are based on presupposition, as the interpretation of one item in a text is dependent on the item that presupposes it.

The presupposition, and the fact that it is resolved, provide cohesion between the two sentences, and in so doing create text (Halliday and Hasan 1976 p4).

Pollach suggests that

pre-supposition makes more of a text than just a collection of isolated words or sentences – it creates cohesion (2003 p30).

Cohesion differs from coherence in that it is a property of the text while coherence is based on the interpretation of the text by the reader. Fawcett (2003 p.91) describes cohesion as the ‘use of grammatical and conceptual devices to guarantee text integrity’ and coherence as ‘the semantic network which glues the parts of a text into a whole’. Halliday and Hasan (1976 p23) differentiate between cohesion and coherence:

The concept of COHESION can therefore be usefully supplemented by that of REGISTER, since the two together effectively define a text. A **text** is a passage of discourse which is coherent in these two regards: it is coherent with respect to the context of situation, and therefore consistent in register; and it is coherent with respect to itself, and therefore cohesive.

In Hallidayan linguistics the relation that is created between two items is referred to as a *cohesive tie*. Halliday and Hasan distinguish five main cohesive devices in English which are the four grammatical elements of *reference*, *substitution*, *ellipsis* and *conjunction* in addition to *lexical cohesion*.

2.1.1.1 Reference

Reference involves items that cannot be interpreted in their own right. It occurs where the reader/hearer identifies a term through reference to another term in close

proximity, thereby setting up a cohesive tie. Common reference elements in English include pronominals, demonstratives and definite articles. Reference can be exophoric or endophoric. With exophoric reference, the information to be retrieved and interpreted is found not in the immediate text but in the context of situation (Halliday and Hasan 1976 p33). In the case of endophoric reference, the referents are found within the text itself. Endophoric reference can be further sub-divided into anaphoric reference and cataphoric reference. Anaphoric reference creates a cohesive link with a preceding item, as in the following example from Baker (1992 p182) where the participant, Poirot, is first introduced by name. There are two anaphora referring back to Poirot, 'He' and 'his':

Hercule Poirot sat on the white sand and looked out across the sparkling blue water. He was carefully dressed in a dandified fashion in white flannels and a large panama hat protected his head.

The following opening line of a short story contains the cataphoric reference 'she', which refers forward to the character 'Nancy':

As she came down the stairs, Nancy glanced at the photograph (Tóibín 2006 p59).

The choice of referential device and the manner and frequency with which these devices are used in a text will generally depend on the text type. A legal text, for example, is more likely to use nominalization than pronominal reference in the interest of clarity. Different approaches to the use of reference are seen not only within the same language but between languages. English tends to use pronominal reference once a participant has been introduced in a text by name.

Hatim and Mason provide an interesting example of how reference can be language and text-specific:

It is a recognised text convention governing the field of discourse of news reporting/investigative journalism in French that a concept referred to in a noun phrase will not be expressed in the same way twice running in a text. Thus, *le dollar américain* will, in a subsequent lexicalisation, become *le billet vert*; *le Président de la République* will become, as well as the anaphoric *il*, perhaps *le chef de l'État* or even *l'Elysée* (1990 p97).

The German pronominal system has a number and gender distinction. Baker suggests that from a cohesion point of view such a system is less constraining as 'different pronouns can be used to refer to different entities within a text with less possibility of confusion' (1992 p182). English, meanwhile, will often revert to full repetition for clarification in the case where a number of participants are mentioned in a particular passage of text.

2.1.1.2 Substitution / Ellipsis

With substitution, one item is replaced by another with no effect on the meaning. In the following example the word 'do' that is spoken by person 'B' is a substitute for 'like ice-cream':

A: I like ice-cream
B: I do too

Ellipsis involves the omission of an item with no effect on the meaning. It is similar to substitution and is referred to as 'substitution by zero' (Halliday and Hasan 1976 p142). The ellided item in the following example is 'bought' (ibid. p187):

Tom bought some sweets and Estella some buns.

Substitution and ellipsis are cohesive devices that are more commonly found in spoken rather than written text and can be nominal, verbal or clausal.

2.1.1.3 Conjunction

The fourth grammatical cohesive device in the Halliday and Hasan model is conjunction which concerns an ‘abstract device between one proposition and another’ (Halliday and Hasan 1976 p13). Conjunction is a cohesive relation that is

‘no longer any kind of a search instruction, but a specification of the way in which what is to follow is systematically connected to what has gone before’ (ibid. p227).

Conjunction is related to the ‘chunking’ of information. This refers, on the one hand to the size or length of each chunk of text and, on the other, to how the relations between the individual chunks of text are perceived and signalled (Baker 1992 p192). Conjunctions can be categorized according to relations, for example, *additive* (and, in addition, furthermore), *adversative* (but, nevertheless), *causal* (therefore, because, thus) and *temporal* (then, an hour later). Smith and Frawley suggest that different genres use different types of conjunction and use them to varying degrees, reflecting the argument and narrative structure of each individual genre (1983 p371).

The colon and semi-colon are potential cohesive devices as they can be used to indicate how different parts of a text relate to each other (Baker 1992 p211).

Conjunctive relations in English can also be expressed in other forms, one example being the verb ‘to follow’ which can be used to suggest temporal relations.

The narrative structure and chunking of information differs not only within the same language from one genre to another but also from one language to another.

Generally speaking, English tends more towards shorter items of information, where the relations between the items are clearly signalled. It also uses a highly-developed

punctuation system to signal these relations (Baker 1992 p193). German on the other hand tends to use complex structures that include a number of subordinate clauses (ibid. p192).

2.1.1.4 Lexical cohesion

Lexical cohesion refers to ‘the cohesive effect achieved by the selection of vocabulary’ in a text (Halliday and Hasan 1976 p274). A lexical item is not cohesive *per se* but enters into a cohesive relation where it ‘involves a kind of choice that is open-ended, the selection of a lexical item that is in some way related to one occurring previously’ (ibid. p318).

Lexical cohesion in the Halliday and Hasan (1976) comprises *reiteration* and *collocation*. Reiteration occurs on the one hand through the ‘repetition of an identical lexical item’ and also through the occurrence of a lexical item that is not identical but that is ‘systematically related to the first one’ (ibid. p284). Reiteration differs from reference in that items that enter into a cohesive relation based on referential cohesion have the same referent. This is not necessarily the case with reiteration. Where a reiterated item has the same referent as the item that is presupposed, it will often be accompanied by a reference item such as a definite article (ibid. p277). In many cases, instances of cohesion are ‘purely lexical, a function simply of the co-occurrence of lexical items’ (ibid. p283).

Reiteration includes *exact repetition* (boy/boy), *synonym* (the boy/the lad), *superordinate* (child/boy) and *near-synonym* (field/pitch).

The second aspect of lexical cohesion in Halliday and Hasan (1976), *collocation* refers to lexical cohesion ‘that is achieved through the association of lexical items that regularly co-occur’ (p274). It means that

any two lexical items having similar patterns of collocation – that is, tending to appear in similar contexts – will generate a cohesive force if they occur in adjacent sentences’ (ibid. p286).

The term has a broad scope, in that it includes, in addition to synonyms, near-synonyms and pairs of opposites, also pairs of words drawn from the same ordered series (for example the occurrence of Tuesday and Thursday in a text will create a cohesive tie) or pairs drawn from unordered lexical sets (ibid. p285).

Each occurrence of a lexical item carries with it its own textual history, a particular collocational environment that has been built up in the course of the creation of the text and that will provide the context within which the item will be incarnated on this particular occasion (Halliday & Hasan 1976 p289).

The importance of collocation in language is expressed by McCarthy (1990 p12):

It is a marriage contract between words, and some words are more firmly married to each other than others. It is an important organizing principle in the vocabulary of any language.

‘Instantial’ or ‘text’ meaning refers to lexical items that co-occur in a specific textual context despite the fact that they may not otherwise co-occur regularly or naturally:

2.1.2 The Hoey model of lexical cohesion

Hoey considers lexical cohesion to be the ‘single most important type of cohesive tie’ (1991 p9) and the ‘dominant mode of creating texture’ where the study of cohesion in text is, to a great extent, the study of patterns of lexis in text (Hoey 1991 p10). Lexis, in particular, fulfils a text-organizing function. In addition to identifying

and categorizing the individual features of cohesion, the Hoey model involves an examination of the extent to which these features contribute to the structure and organization of a text. The model for the analysis of lexical cohesion applies to non-narrative text and reveals both patterns of lexical repetition and lexical networks present in a text.

The lexical items are examined in terms of the presence of repetition links and, where applicable, are categorized according to the type of repetition. These include *simple lexical repetition*, *complex lexical repetition* and *paraphrase*. *Simple lexical repetition* occurs where a word that has already occurred in the text is repeated unaltered, except for a change that might be required from a grammatical point of view, for example 'bear' and 'bears' (ibid. p62). *Complex lexical repetition* occurs when two lexical items share a lexical morpheme but are not formally identical or when they are formally identical but have different grammatical functions, for example 'drug' and 'drugging' (ibid.). *Paraphrase* refers to 'the substitution of one lexical item for another without a loss or gain in specificity and with no discernible change in meaning' (ibid.).

The cohesive relation that comes about between two lexical items is referred to by Hoey as a 'link' and sentences sharing three or more links form 'bonds' (ibid. p91). Hoey differs from Halliday in that he allows for 'degrees of cohesion'. In Hallidayan lexical cohesion a cohesive relation either exists or does not exist. In Hoey, a cohesive link can be strong, becoming a 'bond' or it can be weaker than a bond, remaining a 'link'. This additional categorization allows one to establish which sentences can be considered 'marginal', meaning the sentence in question 'neither

builds lexically upon what has gone before nor provides the lexis for subsequent statements' (Hoey 1991 p105).

2.1.3 Martin

In *English Text* (1992), Martin develops 'analyses which focus on text rather than the clause' (p1). He introduces the concepts of *Homophora* and *Esphora*.

Homophora refers to reference used where 'interlocutors' membership in a particular community means that certain participants can be treated as inherently 'given' (ibid. p122). In the case that the participants are not inherently 'given', reference must be made to the 'context of situation' (ibid.). Esphora refers to forward referencing within the same nominal group (ibid. p123).

2.1.3.1 Conjunction

Martin recognises four main categories of conjunction each of which is sub-divided into additional categories: *additive* (comprises 'addition' and 'alternation': and/or), *comparative* (comprises 'similarity' and 'contrast': likewise/in contrast), *temporal* (comprises 'simultaneous' and 'successive': meanwhile/previously) and *consequential* (comprises condition, consequence, concession and manner: to this end/then/therefore/however/thus).

2.1.3.2 Participant identification

Reference in Martin refers to all the participants realised by nominal groups. He outlines three simultaneous reference systems in participant identification: *generic/specific*; *presenting/presuming* and *comparison/-*. Generic reference is selected where 'the whole of some experiential class of participants is at stake rather

than a specific manifestation of that class', examples being 'deserts, cool deserts' (Martin 1992 p103). Specific reference is the specific manifestation of the generic, for example 'the great Sahara' (ibid.). Presenting reference signals that 'the identity of the participant in question cannot be recovered from the context' (ibid. p102) and is therefore 'strongly associated with first mention' (Martin 1992 p102):

A boy found a frog.

Presuming reference, on the other hand, indicates that the identity of the participant can be recovered from the text. It is generally associated with second mention (ibid.).

A boy found a frog.
He brought it home.

The comparison/- reference system 'makes reference to the identity of the related participants optional' (ibid.).

Martin (ibid. p140) describes a reference chain which is a chain consisting of two items, one presuming and one presumed. In long texts, these chains are extended, 'with presuming items shown to be as well presumed' (ibid. 141). Chains may split and rejoin again (ibid.).

2.1.3.3 Lexical relations

The system for lexical relations consists of three groups: *taxonomic*, *nuclear* and *activity* relations. Taxonomic relations consist of superordinate and composition. Superordination relations are based on sub-classifications (the 'is a' relation', while composition refers to the 'relation of parts to wholes' (the "has a" relation) (Martin

1992 p294-5). Nuclear relations ‘reflect the ways in which actions, people, places, things and qualities configure as activities in activity sequences’ (ibid. p309). Martin suggests that nuclear relations, in the past, have been addressed under the heading ‘collocation’ (ibid.). Finally, activity sequences are related to ‘expectancy relations between activities in activity sequences’ and are based on which nuclear configurations ‘are recurrently sequenced in a given field’ (ibid. p321). The activity sequences in a field are in turn organised into composition taxonomies (ibid.).

2.1.4 Martin and Rose

In *Working with Discourse* (2002) Martin and Rose approach the analysis of discourse not from the point of view of the text as ‘a sequence of clauses’ but of discourse as a social phenomenon that is ‘constructed through text’ (p1). They address three areas that are related to cohesion in a text, namely ‘conjunction’, ‘identification’ and ‘taxonomic relations’.

2.1.4.1 Conjunction

According to Martin and Rose (ibid. p113), conjunction realizes 4 kinds of logical relations: *additive* (and), *comparison* (like), *consequence* (all because) and *time* (then, finally). Conjunction can be used to ‘construe different fields of social activity’, to ‘link events together in a sequence in time’ or to ‘construct the logic of an argument from hypotheses to evidence to conclusions’ (ibid. p111).

2.1.4.2 Identification

Identification involves ‘tracking participants: with introducing people and things into a discourse and keeping track of them once there’ (ibid. p145). The participant can be introduced or ‘presented’ for the first time in a discourse using a proper name

(e.g. Helena) or a lexical item with an indefinite article (A woman) or some other indefinite term (someone). Once a participant has been introduced, s/he or it can then be tracked through the discourse. 'Tracked' is also referred to as 'presumed'. Tracking resources include personal pronouns (he, we), definite articles (the woman), proper names (Helena), comparatives (someone else) and possessive pronouns (my).

Wherever the identity of a participant is presumed, the identity has to be recovered. The identity can be recovered either endophorically (forward or backward reference within the text) or exophorically (outside the text). The use of identification resources differ from genre to genre. Martin and Rose suggest that a narrative or argument text is more likely to display a chaining effect in tracking as opposed to an administrative or legal text which, for reasons of clarity and unambiguity uses forward and backward reference throughout, creating a 'complex lattice of intra-textual relations' (2002 p169).

2.1.4.3 Taxonomic relations

Taxonomic relations construe the 'types of relations that develop between semantic elements as a text unfolds'. They are either 'classificatory' or 'whole-part'. Martin and Rose identify five types of taxonomic relations that can be used cohesively between elements of a text: *classes*, *parts*, *contrasts*, *synonyms* and *repetition*.

Relations of *class to member* could be, for example, 'a make of a car' or 'a breed of dogs'. Typical terms used are 'type', 'category' or 'variety' (ibid. p101).

Relations of *whole to part* can have different labels, depending on the field (e.g. ingredient, component, episode). Relations within a class (Englishman-Afrikaner)

are referred to as co-class. Whole to part is further divided into *facets*, which describe parts that are locations of the whole (the bottom of his soul), and *measures* that refer to a portion of the whole (a cup of coffee, an ounce, a metre) (Martin and Rose 2002 p101). Relations between parts (leg-neck) are referred to as co-parts. *Contrasts* are elements that ‘differ significantly in meaning’. They include elements that are opposed in meaning to each other (win-lose; happy-sad) and also elements that belong to a series of different meanings (hot-warm-tepid-cold) (ibid. p102). Opposed elements, in turn, include ‘antonyms’ and ‘converses’, whereby antonyms come in pairs (happy-sad; win-lose), and converses ‘are associated with converse social roles or locations (victim-perpetrator; give-receive) (ibid.). Series includes scales (pass-credit-distinction-high-distinction) and cycles (Sunday-Monday-Tuesday).

Synonyms ‘are different elements that share similar experiential meanings’ (ibid.). As an example of synonym, Martin and Rose propose the items ‘public hearing’ and ‘open session’ from a Desmond Tutu text from which it is drawn. They suggest that synonym is context-dependent. While the terms above are synonymous in the given context, they may not be synonymous in another.

Finally, *repetition* is described by Martin and Rose as ‘the simplest type of relation between elements’ (ibid. p103). It includes ‘all the forms derived from a single lexical item, for example marry, married, marriage; fail, failing, failure’ (ibid.).

2.1.5 Halliday and Matthiessen

In the 3rd Edition of *An Introduction to Functional Grammar*, (2004) Halliday and Matthiessen divide cohesion into paradigmatic and syntagmatic relations. The

paradigmatic relations are ‘inherent in the organization of lexis as a resource’ and can be interpreted ‘in terms of elaboration and expansion’ (Halliday and Matthiessen 2004 p571). They include *repetition* and *synonymy*. Halliday and Matthiessen (ibid. p577) consider the most direct form of lexical cohesion to be repetition (Algy met a bear. The **bear** was bulgy). Unlike the earlier works of Halliday and Hasan, *hyponymy* and *meronymy* are treated separately to synonymy. The syntagmatic relations, on the other hand, ‘hold between lexical items in a syntagm that tend to occur together, or collocate with one another’ (ibid. p571). Collocation occurs with lexical items that have a ‘tendency to co-occur’ (ibid. p577). The importance of collocation in lexical cohesion is underlined in the comparison of the strength of synonymy and collocation:

Even where there is a relation of synonymy between lexical items, their cohesive effect tends to depend more on collocation (ibid.).

The cohesive effect of synonyms that do not regularly collocate is weakened.

Halliday and Matthiessen also link collocation to the expectations within the text:

Words which are closely associated but without any systematic semantic relationship are nevertheless likely to have a noticeably cohesive effect. This is because collocation is one of the factors on which we build our expectations of what is to come next (ibid.).

Any pair of lexical items can involve both paradigmatic and syntagmatic organisation, given that they ‘represent two different dimensions of patterning’ (ibid.).

The traditional approaches to cohesion outlined above assume a linear, continuous text, which is read in an order that has been pre-ordained by the text producer. It will be demonstrated in section 2.2 that one of the defining features of a web site is that

the reading paths on a web site are not pre-determined by the text producer but, in the main, are created by the user at each reading.

2.2 The web site

The World Wide Web was once described by Berners-Lee (2002) as ‘an abstract imaginary space’. It was to become a powerful means of collaboration and communication between people,

something to which everyone has immediate and intuitive access, and not just to browse, but to create (Berners-Lee 1999 p169)

The Web is, in essence, a network of interconnected documents that can be accessed through the Internet, the physical network of networks created in the early 1970s by Vinton Cerf (ibid. p5-7). These documents, also called web pages, are organised into web sites. Each page of a web site has its own unique address, also called ‘Uniform Resource Locator’ (URL). For example, <http://www.leipzig.de/int/en/> is the address of the English homepage of the official web site for the city of Leipzig. Web pages are created using HyperText Markup Language (HTML) and are displayed on a screen, one at a time, with the help of a browser such as ‘Internet Explorer’. The structure, layout and look of a web page, including the colour, fonts and font sizes are controlled using HTML which also supplements the text content with interactive elements and embedded images (Gramene 2004). The information required here is contained in a text document called the source/HTML document or source code and is expressed in ‘tags’ or ‘metatags’. The HTML document for any given web page

can be accessed when on that page by right-clicking and selecting ‘view source’.

While each web page has its own HTML code, the features of one web page as set out in the HTML code can be copied and inserted into the code for other pages. The replication across a web site of certain elements such as background colour or the position of a navigation bar allows for a high level of consistency, thereby enhancing visual cohesion across the web site.

In the early years of the Web, a web site was considered to be essentially a ‘technical’ tool where a sound knowledge of computer programming was required in order to set up and maintain a web site. Web pages consisted mainly of text as the technology was not yet available to support other elements such as large graphic files or videos. Movement from one page to another came about through hyperlinks.

A ‘hyperlink’ is defined as

a link from a hypertext file to another location or file; typically activated by clicking on a highlighted word or icon at a particular location on the screen (Wordnet Database 2007).

While hyperlinks were an intrinsic part of web sites from the beginning, their true potential had not yet been realised. Text conventions for web sites had not been established to any great extent and as a result, textual content was similar in typeface, length and general layout to printed documents. Cohesion and cohesive devices were treated as for conventional, linear text. A web page resembled more an independent and self-contained unit rather than an interconnected element in the network that is the web site and ultimately the Web itself. The following web page

in Figure 2.1 is a replica of the first web page which was launched by Tim Berners-Lee in 1991:

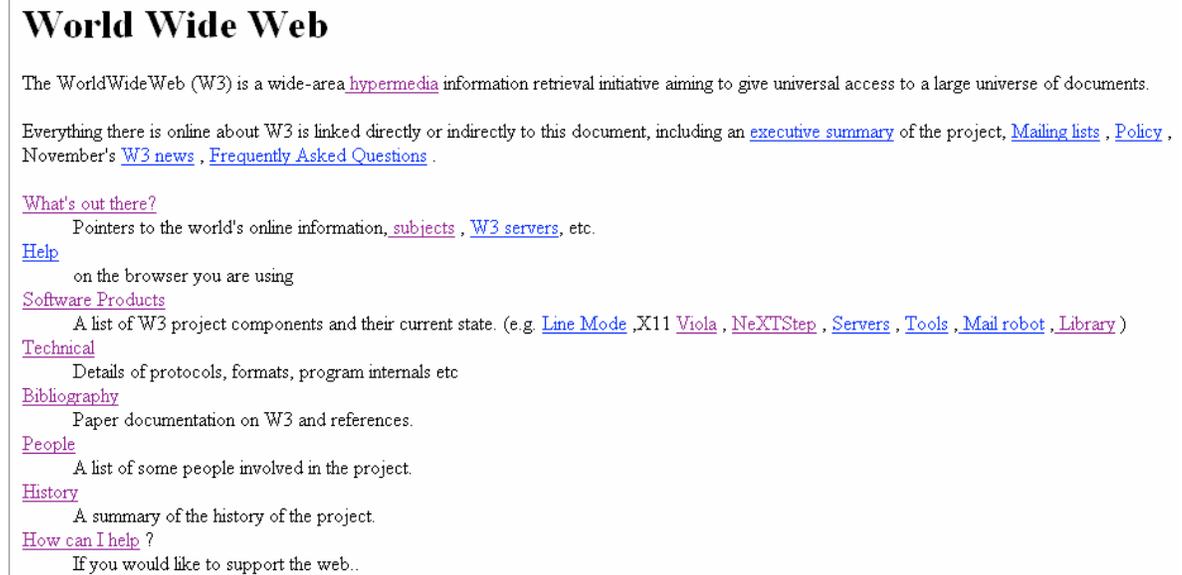


Figure 2.1: The homepage of the first web site, launched by Tim Berners in 1991.
(<http://www.w3.org/History/19921103-hypertext/hypertext/WWW/TheProject.html>)

This site contains text only, in the form of body text (*Everything there is online about W3 ...*) and link text (*What's out there?* or *Subject*). While this web site was undoubtedly revolutionary for its time, on visiting it 15 years later one is struck by the page layout that is similar to the printed medium and the distinct lack of a sense of the overall structure of the site and of the multi-linearity of the site.

Since 1991 the numbers using the Web have increased exponentially, an increase that is due in part to the emergence of Content Management Systems (CMS), which

meant that web design and maintenance were no longer inaccessible to the non-technically-minded.

In the context of a Web site a CMS is a collection of tools designed to allow the creation, modification, organisation and removal of information from a Web site. It is common for a CMS to require users to have no knowledge of HTML in order to create new Web pages (Biz/Ed 2006).

The increase in numbers accessing the Web can be attributed, among other things, to the fact that access is now possible using mobile devices, such as laptops, mobile phones and PDAs (Personal Digital Assistants), and not only via a personal computer. The 'Mobile Web Initiative', a project involving the World Wide Web Consortium (W3C) in collaboration with sponsors such as Nokia and Vodafone, aims to make the Web

reach into places where wires cannot go, to places previously unthinkable (e.g., medical info to mountain rescue scenes) and to accompany everyone as easily as they carry the time in their wristwatches (W3C 2006).

Web sites today have many and varied functions. They can be directory sites (with the content divided by subject area into different categories and sub-categories, one example being the Open Directory Project); blogs (also known as web logs, which are used to post comments on a given subject and for online diaries and discussions); news sites (Sky news online with continuously updated news videos and live tv news feeds), or e-commerce sites for the provision of information on and sale of products and services. The sites can contain multimedia such as podcasts (where the user can listen to a radio programme after the initial broadcast), skype (a peer-to-peer Internet telephony network), music and video downloads (YouTube), social

environments (Myspace and Bebo) and online games (examples include virtual football and chess).

Meaning on the multimodal document, the web site, is created through a combination of language, image, layout and audio content, realized through written and oral language, video, music and graphics, to mention but a few.

Multimodality is not a new phenomenon. However, Kress and van Leeuwen suggest that in Western culture, there has been a distinct preference for monomodality, one example being the ‘most highly valued genres of writing (literary novels, academic treatises, official documents and reports, etc.)’ which were ‘without illustration’ and consisted of ‘dense pages of print’ (ibid. p1). Nonetheless, over the past number of years, various types of ‘multimodal artefact’ have been the subject of analysis: sculpture, architecture and painting (O’Toole 1994); film and television (Iedema 2003); museum (Hofinger and Ventola 2004); printed leaflets, web pages, book pages (Baldry and Thibault 2006). According to Bateman (2008 p7), ‘multimodal written documents are already assuming a central role in many areas of communication’. The term ‘multimodal written document’, also referred to in this case as the ‘multimodal artefact’, refers to a document in which ‘written text’ traditionally played the central role, ‘serving as the principal carrier of information’ (ibid. p1), but where it is now only ‘one strand in a complex presentational form that seamlessly incorporates visual aspects ‘around’, and sometimes even instead of, the text itself’ (ibid.) Examples include patents or legal contracts which generally exist in written form, many of which have illustrations appended. Such documents, web

sites among them, are referred to as ‘pages of spatially orchestrated static content in a combination of modes’ (Bateman 2008 p1), which are delivered by a wide range of technologies, by browsers in the case of the web.

Notwithstanding the multimodality of the web site, the aim of this project is to examine the issue of cohesion on a web site from the point of view of the freelance translator. The focus is, therefore, on the textual elements of a web site and the identification of devices that are capable of functioning cohesively in an online environment. The importance of language on the Web is highlighted by Nielsen and Loranger who maintain that the user visits a web site in order to access web content which is, to a great extent, expressed through language. Language therefore should be the focal point of a web site (2006 p 254). According to Ravichandran et al, as the Web is increasingly viewed from a social perspective, so the role of language (a social phenomenon) becomes central (2004 p3). This is a view supported by Crystal (2001 p viii):

... notwithstanding the remarkable technological achievements and the visual panache of screen presentation, what is immediately obvious when engaging in any of the Internet’s functions is its linguistic character. If the Internet is a revolution, therefore, it is likely to be a linguistic revolution.

The findings of the Eyetrack III research project (Outing and Rual 2004) suggest that ‘text rules on the PC screen – both in order viewed (in relation to images) and in overall time spent looking at it’. It is, therefore, reasonable to suggest that the central feature of many web sites is verbal language. Kress and van Leeuwen note that ‘Writing ... is freeing itself from its previously seemingly fixed appearance on the

page: in a line, followed by other lines, moving in direction from left to right, down the page ... directionality is now a variable, a matter of choice, in many genres' (Kress and van Leeuwen 2001 p125). Nowhere is this more evident than on a web site.

2.2.1 Web sites and hypertext

The text on a web site is present in many different forms and includes, firstly, text that is visible on the screen, examples of which are seen in Figure 2.2 below:

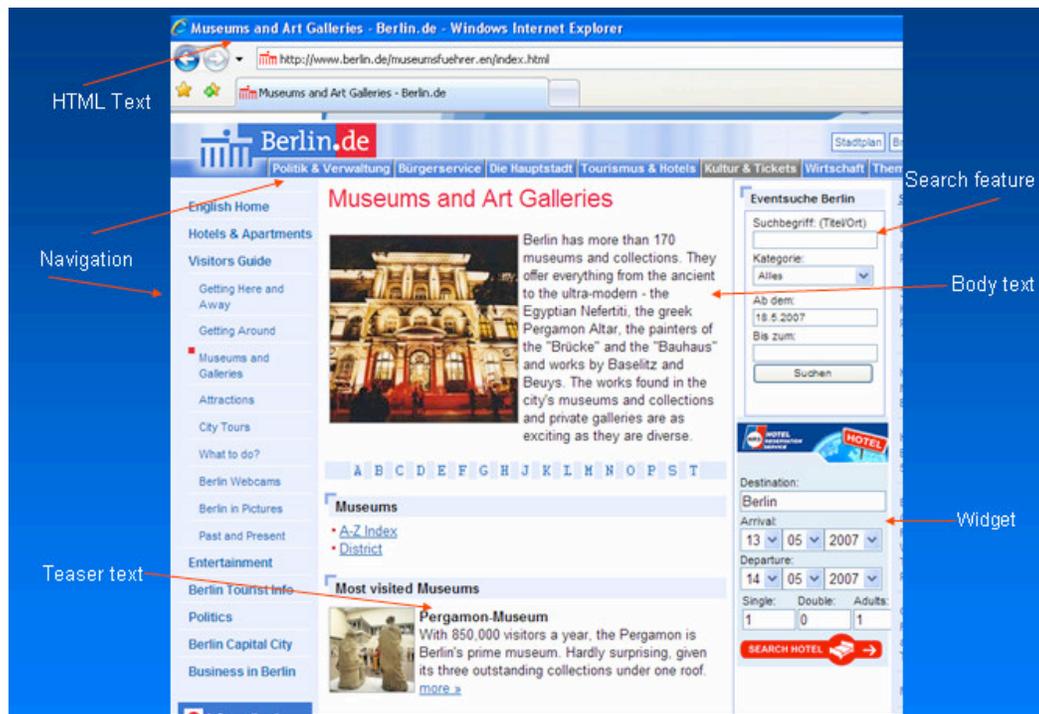


Figure 2.2: Examples of text on English homepage of Berlin web site (www.berlin.de/museumsfuehrer.en/index.html)

This example includes navigation text (navigation bar: 'Politik & Verwaltung' and menu item: 'Getting here and away'), teaser text (Pergamon-Museum), body text

(‘Berlin has more than 170 museums and collections. ...’) and widgets (‘Eventsuche Berlin’, ‘Hotel reservation service’).

Of significance for user interaction with web content is also HTML text, in particular the page title, description content metatag and keywords metatag and, where present, the h1 header text:

```
<title>muenchen.de - Accommodation, Hotels in Munich</title>
  <meta name="description" content="Information on hotels and
accomodations in Munich" />
  <meta name="keywords" content="Hotel Munich, Munich Hotel, hostel,
private accomodation, hotel, camping, Accomodation Munich" />
```

Figure 2.3: Title tag, description content metatag and keywords metatag: English Accommodation page, Munich site

(http://www.muenchen.de/Tourismus/Accommodation_Hotels/7604/index.html)

The content of the title tag, <title>, seen in Figure 2.3 above, ‘muenchen.de – Accommodation, Hotels in Munich’, is referred to as the ‘page title’. The page title for the current web page, if defined, is shown at the top of the screen on the blue reverse bar. The h1 header text appears directly on the web page. The description content and keywords metatags do not appear on the web page.

A third element of text that is not placed on a web site but that nonetheless plays an important role in decisions on terminology for the translated web site is the keyword or keywords used by a web user when searching for specific content on a web site using a search engine.

Figure 2.4 shows a search query on Google using the keywords ‘hotels in munich’:



Figure 2.4: Google Ireland search feature
(<http://www.google.ie/>)

Figure 2.5 below is one of the search results returned by the search engine in reply to the above search query:

[muenchen.de - Accommodation, Hotels in Munich](#)
Information on **hotels** and accomodations in **Munich**.
www.muenchen.de/Tourismus/Accommodation_Hotels/7604/index.html - 76k -
[Cached](#) - [Similar pages](#)

Figure 2.5: Google search result for search query ‘hotels in munich’.
(<http://www.google.ie/search?q=hotels+in+munich&hl=en&start=10&sa=N>)

The importance of a search query lies in the fact that it is the language and terminology used in an online search by the user, which will ideally inform decisions on terminology made by the translator as part of the overall cohesion of a web site.

The defining feature of the WWW is, however, hypertext. The concept of hypertext

is one of the basic principles underlining the general objective of the Web, which encompassed

the decentralised, organic growth of ideas, technology, and society. (It is about anything being potentially connected with anything (Berners-Lee 1999 p1).

Hypertext consists of two main components, firstly the items of information that are linked to each other and secondly the actual links between those chunks. Hypertext includes 'hypermedia' which describes the linking to each other of different media, for example text with video or audio information (Rosenfeld and Morville 2002 p73).

Baldry and Thibault (2006 p156) suggest that hypertext is not a complete departure from all that has gone before. It is rather

a hybrid of precursor genres such as verbal text, visual images, and multimodal combinations of these, on the one hand, and the new meaning-making possibilities of the virtual environment of hypertext, on the other.

Hypertext has created a true network of the Web and one where all documents are linked directly or indirectly to all others on a web site and across the entire Web.

Hypertext is a means of organizing material that attempts to overcome the limitations of traditional text and in particular its linearity. There is no one beginning or end on a web site, as is the case for example with a book. Landow (2006 p110) writes that the 'concepts (and experiences) of beginning and ending imply linearity' and that, assuming that hypertextuality 'has multiple sequences rather than an absence of linearity and sequence', it 'provides multiple beginnings and endings

rather than single ones'. Information 'chunks' are linked to each other

by multiple paths, chains, or trails in an open-ended web – it is multisequential or multilinear rather than non-linear (NUS 2006).

The user can enter a web site on any page, be it the homepage or the most obscure page. It is suggested that, through hypertext, information is available to the user

in much the same way that humans think – by association rather than by linear sequence (Amberton 2007)

Despite the apparent lack of structure with hypertext, where the '*locus of control*' shifts to the user, Baldry and Thibault suggest that hypertext and activities related to it are anything but unstructured. Hypertext underlines the fact that there are no 'a priori structures that cause or guide meaning-making activity from start to finish' (2007 p118). Rather, there is a

multiple, parallel, open-ended, backlooping interplay of texts, genres, semiotic modalities, technologies and the user's perceptions and actions'

In effect, the Web user decides how, what and in which order s/he wishes to read or access the information available on web sites. Baldry and Thibault (ibid. p116) refer to the 'hypertextual trajectory', a trajectory referring in particular to the

meaning-making pathways that are created when users of websites create links from one web page to another, from one website to another, and so on, as they navigate or author their way through a website or from one website to another

As an analytical tool, the trajectory is used to map the user path through a web site and to 'reconstruct users' pathways as a form of multimodal text' (ibid.).

2.2.2 Analysing multimodal documents

As previously mentioned, this research is concerned primarily with the analysis of text elements on a web site. Approaches by Schriver to the analysis of multimodal documents deal with design elements in the printed media, however, certain aspects are applicable for the purposes of this research on the online environment. Schriver suggests that design features are often carried by text elements (1997 p342), of which there are two kinds: 'simplex' and 'complex'. 'Simplex' elements 'often depend on their genre (e.g., reports have executive summaries, online help has procedures ...)' and are individual elements of page design (ibid. 343). However, there are groups of such individual elements that Schriver calls '*rhetorical clusters*'.

By *rhetorical cluster* I mean a group of text elements designed to work together as a functional unit within a document. Rhetorical clusters act as reader-oriented modules of purposeful and related content. They are comprised of visual and/or verbal elements that need to be grouped (or put in proximal relation) because together they help the reader interpret the content in a certain way (ibid.).

Examples of rhetorical clusters are headings and sub-headings. The advantage of considering a document from the point of view of clusters, is that the document can be thought of as a

field of interacting rhetorical clusters. If the document is well designed, the clusters orchestrate a web of converging meanings which enable the reader to form a coherent and consistent idea of the content (ibid. p344)

The identification of such rhetorical clusters on a page enables the designer to see the text from the user/reader point of view, which underpins Schriver's

recommendation to approach document design in a user-oriented manner. For the analysis of multimodal documents Baldry and Thibault provide a transcription of the elements under investigation. In order to identify those elements that will be the subject of transcription, they employ, similar to Schriver, the concept of the 'cluster':

Our use of the term cluster refers to a local grouping of items, in particular on a printed or web-page (but also other texts such as manuscripts, paintings and films). The items in a particular cluster may be visual, verbal and so on and are spatially proximate thereby defining a specific region or subregion of the page as a whole. The items in a cluster are functionally related both to each other and to the whole to which they belong as parts (Baldry and Thibault 2006 p31).

The Baldry and Thibault approach is adopted in order to define the elements of the corpus web sites that will be examined in terms of cohesion. Returning to Figure 2.2 which displays various instances of text on a web site, we immediately recognise a number of 'clusters' and therefore potential areas for the investigation of cohesion.

Figure 2.6 below shows individual clusters which are marked out and numbered 1 to 5:

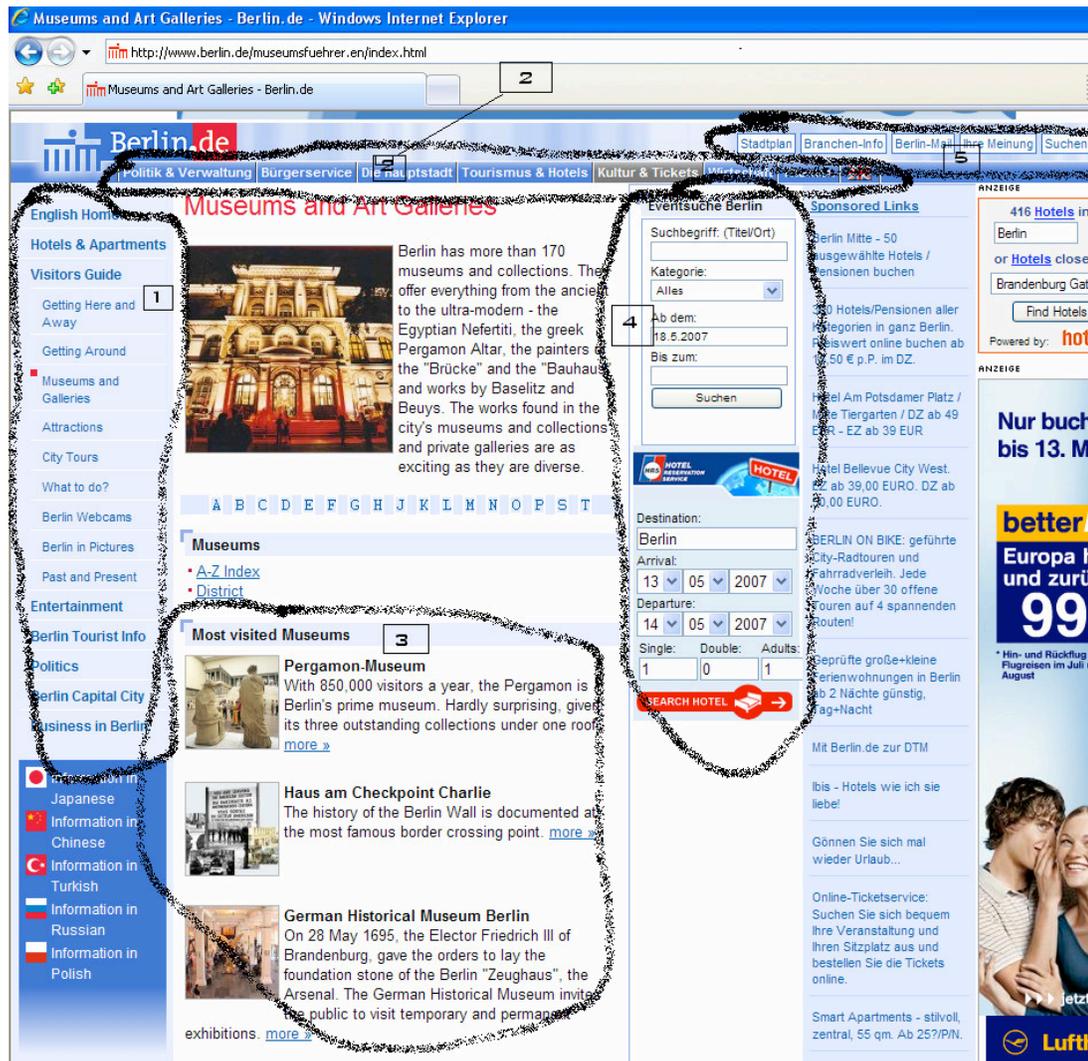


Figure 2.6: Examples of text clusters on English homepage of Berlin web site (www.berlin.de/museumsfuehrer.en/index.html)

In keeping with the Baldry and Thibault (2006) use of the term ‘cluster’, the members of the clusters outlined in Figure 2.6 are both visual and verbal and

spatially proximate. The items of each cluster are functionally related to each other and to the web site as a whole.

The clusters in Figure 2.5 and their functions can be described as follows:

Number 1 is a menu bar and is navigational (local). It has the function of providing a link to specific sub-sections of the site. It is functionally related to number 2 which is a navigation bar and therefore navigational (global). It has links to the main sections or subject areas of the site. Numbers 1 and 2 are related to number 5 which consists of a group of hyperlinks and is therefore navigational, linking to diverse types of information. Number 3 is a teaser text, indicating the presence of additional information in a longer form than is the case with a one-term link. It is related to number 4, a widget, as both ultimately indicate additional information

A sixth cluster of similar items on the page could be the icons at the bottom left of the page which would be referred to as ‘navigation – access to TL content’.

In addition to the related clusters of ‘navigation’ (menu bar and navigation bar etc.) and ‘content’ (widget and teaser text), a third ‘cluster’ will form part of this project, given its significance in terms of the user web experience and therefore its importance in cohesion on a web site. This cluster comprises the relevant HTML tags, title tag and description content and keywords metatag, most of which are not visible on the web page but are nonetheless significant.

Together, these groups of items represent three fundamental differences in the way the web user engages with a web site compared with how s/he deals with conventional text. The first change concerns the fact of finding and accessing Web content using a search engine (directly related to the HTML tags and metatags).

Second is the manner in which the user can navigate the pages of a web site and between web sites and retain a sense of orientation. The third change relates to how the user engages with various on-page elements of a web site. On the basis of these three areas, this dissertation will provide evidence of the need to re-examine the nature of cohesion on a web site and to identify those elements on a web site that are capable of functioning cohesively.

2.2.2.1 Finding web content using a search engine

The most popular means of accessing Web content in Ireland (Amárach Consulting 2008) and abroad is with a search engine such as Google. This type of content search is very much user-driven as the user formulates the search query for input into a search engine and can re-formulate it at will, if the results are not at first satisfactory. The user then has immediate access to a list of web sites that contain content on the subject area required and can decide whether to access one of the search results or not or whether to carry out a new search.

In terms of online search, it can be said that the translated web site is functioning cohesively where the web user succeeds in finding and accessing content that is relevant to his/her requirements, and valuable and is in the language required. The translator has an important role to play when defining TL keywords as decisions in terms of 'equivalence' are no longer sufficient in themselves in an online environment. The use of the term 'equivalence' in this study is similar to its use in Hervey, Loughridge and Higgins (2006 p15), where it refers to a translation methodology that avoids 'an absolutist attempt at *maximizing sameness* in things that are crucially different (ST and TT), in favour of a relativist attempt at

minimizing dissimilarities between things that are clearly understood to be different’.

Equivalence in the context of web site translation of necessity includes a choice of terminology that reflects ‘user language’, which is the language used by the web user when searching for content on the Web.

It is not sufficient for the SL content on a multilingual site to be optimised for search. The TT must also be treated as web content in its own right and optimised in its own right otherwise it runs the risk of simply not being found by the user in the given subject area and language as it may not be indexed correctly by search engines.

2.2.2.2 Web site navigation

The second innovation brought about through hypertext involves navigation and concerns how cohesion facilitates navigation on a web site and provides context and a sense of orientation on the page on which s/he enters a web site, regardless of which page it is. The homepage of a web site is often considered to be the ‘main’ or ‘entry’ page and the page to which the user is often directed initially. However, studies show that ‘interior pages’, that is pages other than the homepage, now account for 60% of initial page views (Nielsen and Loranger 2006 p29).

The practice of guiding the user directly to an interior page on a web site is called ‘deep linking’ (Nielsen 2002a). Deep links are more likely to satisfy user requirements as the user is taken directly to pages that provide content on specific issues. The user does not have to first negotiate his/her way from the home page through additional pages and sections to the required destination page (Nielsen and Loranger 2006 p29). However, deep linking is comparable with a book reader

opening a novel in the middle of Chapter V and attempting to make sense of the content based on the information provided on that one page. The difference between the book and the web site is that the book reader is unlikely to be able to make any sense of the page s/he has arrived at without first knowing what has gone before in the novel. Regardless of the page of entry or level in the web site hierarchy, the user should be in a position to immediately recognize his/her location in terms of the overall content structure of the site and make an informed decision as to where to go from there. In addition, studies show that the user spends an average of only 27 seconds on an interior page after arrival which underlines the need for quick recognition of where s/he can go from there (Nielsen and Loranger 2006 p35).

A web owner has no control over incoming links to his/her site. An incoming link is a link placed on a web site to a page on another site. It is, for example, possible that the user accessing content on a multilingual site through a link on another site could be guided to content in a language other than the one indicated in a link and expected. In such a case, the global navigation system of the site must be capable of guiding the user to content in the required language. It is not uncommon for a user to access the same pages of a particular web site each time using a different path to previous visits. In effect, new cohesive patterns are created with each repeat visit to a web site.

2.2.2.3 Interaction with on-page elements

Eyetracker studies are used to examine user activity and reading patterns on a web page. By tracking the eye movements of the user when accessing web sites specific information can be recorded, for example the areas of a page that are viewed, length

of time spent viewing certain areas, position and number of clicks and how far down the page the user scrolled, if at all. This information is collated and reproduced visually on a *heatmap*, where different colours are used to indicate the different levels of activity at different points of a web page. The following heatmap was generated as part of the Eyetrack III project which examined user behaviour on mock news web sites and real multimedia (Outing and Ruel 2004). Figure 2.7 is a record of user behaviour around a teaser text:



Figure 2.7: Heatmap from Eyetrack III, showing user behaviour around a teaser text (<http://www.poynterextra.org/eyetrack2004/main.htm>)

The orange area is the area most viewed and the dark blue the area that is least viewed.

The Eyetrack III study returned the following key findings:

- users tend to scan a page first before deciding to read
- dominant headlines most often draw the eye of the user on entering a page and users read the first few words of a headline only
- in the case of a headline followed by a blurb or teaser text, the user will read the left one-third of the teaser text, if s/he reads beyond the headline at all.

The study also suggests that larger font sizes promote scanning while smaller font sizes promote reading. While it could be argued that the material used in the Eyetrack project - online newspapers - is intended less for in-depth reading, the

results nonetheless indicate that *skimming* and *scanning* are typical user behaviour on a web site. The technique *skimming* is used as a quick method of identifying the salient points of a text and is said to be almost four times faster than the normal reading speed (4 2 Explore 2005). *Scanning* involves moving one's eyes down a page quickly, seeking specific words and phrases. Nielsen and Loranger suggest that users will always scan to get an overview of the content on a page and, depending on the level of information required, will sometimes read. They see the Web as a 'user-directed medium' (Nielsen and Loranger 2006 p258) where scanning is

an efficient method to home in on useful content. It takes less cognitive effort, so users can focus attention on fruitful areas (ibid. p259).

In addition, given the level of information available on any given topic, it is suggested that the user must be in a position to scan and prioritize when seeking information.

Nielsen and Loranger differentiate between 'content' pages and 'navigation' pages.

A 'content' page consists of more in-depth information on a certain subject (ibid. p258) and will generally consist of continuous text. A 'navigation' page, on the other hand is, as the name suggests, a means of moving closer to a 'content' page which contains the information that the user is attempting to access on a site. User behaviour differs on a page depending on whether it is a navigation or content page. A navigation page, which can contain teaser texts and lists of links in addition to a small amount of body text, is more likely to be scanned and skimmed than read word for word. It is suggested that users do not spend as much time on a content page as

they do on a navigation page as they feel more productive when they are clicking links and moving from one topic to the next (Nielsen and Loranger 2006 p269).

Usability tests carried out by Nielsen and Loranger make a strong argument for the structuring of continuous text on a web page into small units which improves cohesion and therefore readability:

Short paragraphs surrounded by white space appear more approachable than a solid wall of text. Information shortened into digestible pieces facilitates scanning by allowing people to see natural breaks and absorb the information in manageable pieces (2006 p282).

The use of the 'one topic and one idea' sentence is recommended (ibid.) as is the use of the 'inverted pyramid' for a content page on a web site. The inverted pyramid refers to a text where the conclusion is provided at the beginning and the content is fleshed out throughout the rest of the text (ibid. p269). Cohesive reading patterns in the multilingual web site that support scanning and skimming are lexis-based, relying on consistency and positioning of terminology on a web page.

2.3 The relevance of traditional models of cohesion to a web site

Traditional models of cohesion as outlined above can, in general, be used for the analysis of cohesion of continuous, body text on a web site. However, it must be borne in mind that continuous text on a web site is not similar in all respects to conventional text in the print media, though this would appear to be the case. As with all instances of text on a web site, continuous text is part of a network and should be seen, not as a text in isolation but one that is linked lexically and, sometimes through hyperlinks, to the surrounding text and images on the same page,

on other pages on the site or even to other web sites. It must also be taken into consideration that the continuous text may have been optimized for search engines. In this case it may display patterns and devices of cohesion that are unlike those that would normally be expected in this type of text in the printed medium. One such example of an optimised continuous text is the presence of an unusually high level of lexical repetition as opposed to referential cohesive devices, particularly in the first few paragraphs of the text. Lexical repetition could be a reflection of the need in terms of a search engine to include a certain number of the keywords for the page in the first hundred words of the text. In addition, the continuous text may be prepared for online reading by chunking it in accordance with web site usability guidelines, one of which would involve the inclusion of 'one-idea' paragraphs on a web page (Nielsen and Loranger 2006 p275). This means that the continuous text might differ from the standard chunking of text that is common to a similar text type but in the print medium. Continuous text is not examined for cohesion in this project.

Traditional models of cohesion apply to linear text where both referential cohesive devices and lexical chains are easily traced forwards and backwards, beyond the sentence and even the page.

However, the text that is the subject of analysis in this project is non-linear/multi-linear, is for the most part hypertextual in nature and is relatively short (page title, teaser text heading), at times consisting only of one term (a keyword). These items of text are not part of a continuous text. In addition, they appear not only on the visible web page, where they are positioned strategically, but are also significant in

search results and in connection with access to other pages on the same site and to other web sites.

According to Martin and Rose (2003 p173):

In 'little texts' such as headlines, telegrams, SMS messages on mobile phones, titles, labels, diagrams, billboards and so on, determiners are more often than not left out, so the distinction between presenting and presuming is neutralized.

It is reasonable to assume that this also applies to web site text such as that under investigation, particularly in view of the fact that a web site can be entered on any page, regardless of whether it is the homepage or any other. Reference as a cohesive device is unsuitable for such web site text as a means of creating cohesive relations between web pages as the 'hypertextual trajectory' will not always be the same for all users or even for the same user on a second accessing of the same content. On a web site, a reading path can unfold across a number of web pages and can also involve leaving the web site temporarily to access a different site and then returning to the original site. There is no pre-ordained, linear pathway. For example, 'He' or 'It' cannot function cohesively unless the person or thing with which a cohesive link is intended is in the immediate vicinity on the same web page. Equally with a lexical cohesive resource such as 'On the following page', a cohesive link can only be created to the page in question, wherever that is, if a link is provided immediately after the text or if the text 'On the following page' is the link.

The cohesive device of conjunction, which helps to chunk text and to indicate the direction of an argument, is not viable as a cohesive device for the text elements

under investigation, as these elements are not part of a continuous text requiring the linking of 'chunks' or the development of an argument.

Some elements of lexical cohesion examined above, for example exact repetition, simple repetition, superordinate/hyponymy and of taxonomic relations (classes, parts repetition) are relevant to cohesion on a web site.

Where a keyword is defined for a web page, reflecting the content of that page, and is indexed by a search engine, a cohesive tie is created between the user, the search result and the destination web site where that same keyword is used by the user in online search. A keyword that is used as a link label and that links to a page with a page title or a heading that includes the same keyword, creates an anaphoric link to that page through the cohesive device of repetition.

Synonymy, on the other hand, is problematic in terms of cohesion on a web site, given that cohesion is created through the exact repetition of keywords both on a web page and in HTML tags. The web spider can match a keyword in repetition but will not recognise the connection between synonyms. It is precisely the repetition of naturally-occurring keywords on a web site that creates a cohesive network not only for the user, in terms of scanning and skimming, but also for a search engine in terms of page ranking based on a network of terms in repetition.

Superordinate/hyponym are relevant in that the hypertextual trajectory frequently involves movement from the general towards more specific content. Baldry and Thibault (2006 p113) suggest that the web site has generic features and that the homepage, for example, is a type of functional component within the greater structure of the web site and itself has the characteristics of a superordinate genre

with other mini-genres. Returning to our example in Figure 2.5 above, a search for 'hotels in munich' could take the following path if that link is used to access the site: the section/page 'Accommodation, Hotels' has the sub-sections 'Hotel Reservations', 'Recommended Hotels', 'Private accommodation, Apartments' and 'Youth Hostels, Camping'. The term 'Accommodation' is a superordinate for 'Apartments', 'Youth Hostels' etc.

If collocation is assumed to occur with lexical items that have a 'tendency to co-occur' (Halliday and Matthiessen 2004 p577), then a concept of collocation in an online environment can be proposed. This occurs where two or more keywords have a tendency to co-occur in a specific context, for example in online search.

Appropriate and corresponding co-occurrence can be established in the TL and the TT prepared in such a way as to accommodate such occurrences.

It has been shown that certain cohesive devices (repetition, superordinate/hyponymy etc.) relevant to conventional text can function cohesively in an online environment. In particular, repetition is significant for cohesion in Online Search, in order to match terminology on a web page with the user language in a search query, via a search engine. In the case of Navigation, repetition is also significant, for example where a keyword in a page heading is in a cohesive relation with the link label via which the user arrived at that page. Lexical cohesion based on a superordinate relation is also relevant in Navigation, whereby section labels are seen to be superordinate to their sub-section labels (e.g. 'Sightseeing' to 'museums').

However, lexical cohesion alone is not sufficient to explain how cohesive links are created in a web site environment, given the multilinearity of reading pathways.

Based on the web site as a ‘visual-spatial unit’ (Baldry and Thibault 2006 p104), the relevance of aspects of visual composition for web site cohesion will be examined.

2.4 Cohesion and visual composition

Cohesion is a term that is no longer considered solely in the context of verbal text but is also understood to be a significant aspect of visual composition. The model for the analysis of visual composition by Kress and van Leeuwen first introduced in 1996 provides the basis in this study for the examination of cohesion in a visual context. The Kress and van Leeuwen model draws on Hallidayan linguistics and is based on a definition of text as ‘any instance of communication in any mode or in any combination of modes’ (Kress 2003 p48).

All elements of a visual composition are considered to be ‘text’. Verbal text is viewed as a ‘visual block’ and not as a linguistic item. Communication in the visual composition is realised through three metafunctions, the *representational*, the *interactive* and the *compositional* which reflect the *ideational*, the *interpersonal* and the *textual* metafunctions in Hallidayan linguistics.

The *representational* metafunction represents the

interactions and conceptual relations between the people, places and things depicted in images (Kress and van Leeuwen 1996 p119).

The participants in a visual composition are referred to as ‘representational participants’ (ibid. p46). The *interactive* metafunction represents the relation

between the represented participants and the viewer of the visual composition. It involves concepts such as ‘contact’, ‘social distance’ and ‘attitude’ (Kress and van Leeuwen 1996 p154).

The *compositional* metafunction denotes

the way in which the representational and interactive elements are made to relate to each other, the way they are integrated into a meaningful whole (ibid. p181).

The metafunction *composition* consists of three elements, *framing*, *information value* and *salience*. While the concept of ‘framing’ closely reflects cohesion in a linguistic context, it will be demonstrated in this dissertation that the concepts of ‘information value’ and ‘salience’ also play a vital role in the creation of textual cohesion in the translated web site.

2.4.1 Framing

Framing refers to the connectedness or disconnectedness to each other of elements of a visual composition. Similar to the Hoey model of lexical cohesion, the Kress and van Leeuwen model allows for varying levels of connectedness or disconnectedness:

The more the elements of the spatial composition are connected, the more they are presented as one unit of information, as belonging together (ibid. p215).

Connectedness is achieved through harmonizing colours and the absence of frames or empty spaces between elements of the composition. The degree of connectedness depends on the thickness of the lines or frames and the changeover from one colour to another. Connectedness in some visual compositions is also created through the

use of vectors. A vector is a real or imaginary line joining two elements. It is created either by the shape of an object, or the gaze of a person towards another person or object. The vector begins at the most salient object, which is the object that has first drawn the attention of the viewer. It then draws the viewer's gaze towards the object at the end of the vector (Kress and van Leeuwen 1996 p216). Disconnection in a visual composition is achieved through contrasting colours, empty spaces and the presence of lines and frames (ibid. p214). Figure 2.8 shows the main navigation bar from the official Frankfurt site with the tab labels 'Leben in Frankfurt', 'Tourismus', 'Kultur', 'Wirtschaft' and 'Rathaus':

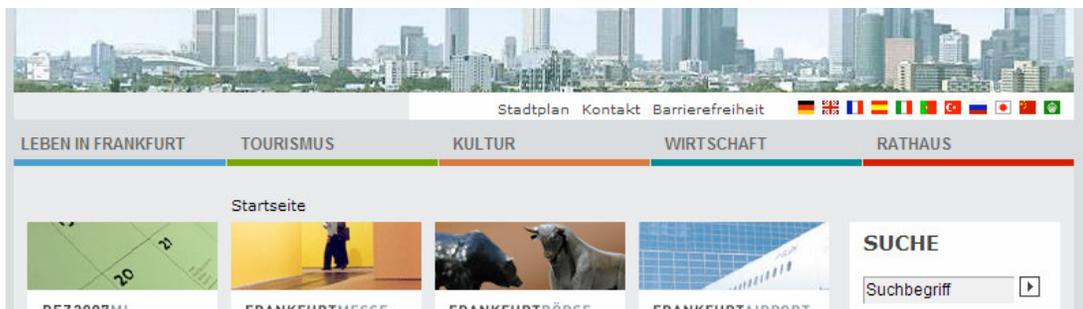


Figure 2.8: Main navigation bar, Frankfurt site
(<http://www.frankfurt.de>)

This navigation bar demonstrates, on the one hand the connectedness to each other of the navigation tab labels. The connectedness is created by the lack of lines between the tabs, by their common (black) colour and the common (grey) background colour and a lack of framing around each label. It could also be said, however, that there is a certain disconnect between the individual labels of the navigation bar due to the quite considerable space between each label. In such an instance a lack of connectedness appears to be more advantageous for the textual

cohesion of the site as it sets each label apart from the other and in so doing supports scanning. This enables the user to see the tab labels for the different main sections of the web site immediately.

At the same time, the level of connectedness can have a positive effect on overall cohesion, for example in the case of the teaser text. Eyetracker III suggests, for example, that if there is a significant disconnect or gap between a heading and its accompanying text, the user might only read the heading and will not bother reading the accompanying text (Outing and Rual 2004). The following two examples are teaser texts from the German and the English content of the Munich site respectively:



Figure 2.9: Teaser text, Munich web site
(<http://www.muenchen.de/>)



Figure 2.10: Teaser text, Munich web site
(<http://www.muenchen.de/home/60093/>)

While the body text (Die Symphoniker ...) in the German example follows directly after the heading (Feier zum Nikolaus) without a space, this is not the case with the English text (Facts and figures ...) and its preceding heading (City Hall), between which there is an empty space denoting a disconnect. Regardless of how slight such a gap may be, it could still affect the cohesion of the textual items.

2.4.2 Information value

Information value refers to the actual positioning of individual elements in the overall visual composition:

... the role of any particular element in the whole will depend on whether it is placed on the left or on the right, in the centre or the margin, or in the upper or the lower part of the picture space or page (Jewitt and Oyama 2001 p147).

The space can be divided horizontally into left/given and right/new and top/ideal and bottom/real. The 'Given' is considered to be something that is 'a familiar and agreed departure point for the message' and is commonsensical and self-evident. The 'New'

on the other hand is viewed as something unfamiliar to which the viewer or reader 'must pay special attention' (Jewitt and Oyama 2001 p148). The centre space is

seen as what holds the 'marginal' elements together – (they are).. belonging to it, subservient to it...' (ibid. p149).

Cohesion on a web site is enhanced by the consistency created through the placing of the same items in the same position on all pages of a web site. Such items are, for example, the main navigation bar on a site or the link on all pages of a site to the homepage.

Such consistency across a web site aids cohesion in that the user knows exactly where to find certain items when moving from one page of the site to another.

Nielsen refers to such consistency as 'one of the most powerful usability principles' (1999). This is based on the idea that if the user knows what to expect, s/he feels in control of a particular system or web site and is therefore generally more comfortable with such a system.

A word order that places the most important information as far left as possible in a line of text, which is in the 'Given' position, facilitates cohesion and therefore enables the user to quickly identify the salient points of a text through skimming.

This is in keeping with the finding of the Eyetrack III report with regard to teaser texts which suggests that the user will often read only the left one-third of the body text of a teaser.

2.4.3 Saliience

‘Saliience’ occurs where certain elements of a visual composition are made to appear more prominent than others, where

the elements ... are made to attract the viewer’s attention to different degrees, as realized by such factors as placement in the foreground or background, relative size, contrasts in tonal value (or colour), differences in sharpness, etc. (Kress and van Leeuwen 1996 p183).

On a web site, saliience is achieved by varying the size of the images depicted, through the use of different colours, typefaces and typeface sizes and italics and bold print. An example of saliience created through bolding (headings) and colour (red icons) is seen in Figure 2.11 below from the German homepage of the Berlin site:



Figure 2.11: Saliience markers – bolding and colour
(www.berlin.de)

In this example, the headings (‘Glückwünsche für Ertl’ and ‘Bühne: “Rhythmus Berlin”’ etc.) are more prominent than the navigation labels (‘Politik & Verwaltung’, ‘Bürgerservice’ etc.). This is due to the fact that, on a web page where blue is the dominant colour, bold black against a light blue background is more salient than white against a dark blue background.

Salience on a web site enhances or supports lexical cohesion on two levels. Not only is a potentially cohesive text created for the user in that scanning is facilitated through bold or italicised keywords which enables the user to move swiftly through the text, from one key item to the next. Important information is also provided to search engines that give more weight to lexis that is bolded, italicised or where keywords are at the same time hyperlinks.

Textual cohesion on the translated web site relies not only on the lexical features of a text such as reiteration or simple repetition, but as has been shown, to a great extent on aspects of visual composition, namely framing, information value and salience. It can be said that lexis in cohesion on a web site follows not only the logic of time but also the logic of space.

2.5 A definition of web-specific cohesion

Given the considerable differences in means of access and reading patterns of web content compared with conventional text, it can be said that traditional models of cohesion are not sufficient to account, in their entirety, for cohesion on a web site. Therefore, a definition of web-specific cohesion is proposed, which refers primarily to those instances of text on a web site that do not lend themselves well to analysis using traditional models of cohesion:

Web-specific cohesion refers to the network of lexical items, referred to as specifically-defined keywords that link various parts of a web site in a multi-linear fashion. These lexical items are not cohesive *per se* but rely on aspects of visual composition in order to function cohesively.

Keywords provide a cohesive link mainly through 'exact repetition'. They reflect the terminology used or likely to be used by the user in a specific situation and a specific subject area. Groups of keywords are referred to as keyword phrases. They are collocational in nature as they reflect the multi-word keyword search that are becoming increasingly common as opposed to single keywords (Nielsen and Loranger 2006 p167) and may only appear in this combination and word order in a very specific environment.

It is important that considerations of cohesion on a web site are addressed not only in terms of a single page or even a single web site. Given the interconnected nature of the pages and sites on the Web, decisions made with regard to a page or one web site will inevitably impinge on the level of cohesion of other sites. One example is the provision of a link on a translated site to another web site where, instead of linking to content in the same language, the link is to content in an entirely different language. A cohesive link between keywords in repetition is therefore not possible. The textual items that contain the cohesion-creating text include HTML text (page title tag and description content metatag), hyperlinks (navigation bar, links) and elements of page content (teaser text, and body text).

Cohesion on a web site is not simply an issue for the SL content but must also be regarded as an integral part of a well-functioning TL content. The appropriate use and positioning of keywords and the inclusion of TL content where required are important factors in creating cohesion in the translated web site.

2.6 Approaches to the translation of web sites

There are four principal ways in which multilingual content is provided for a web site: Machine Translation (MT), web site localization, by commissioning a translation company or by directly commissioning a freelance translator.

2.6.1 Machine translation

MT involves the automated translation of a document without the input of a human translator. It is frequently used for the translation of web sites with extensive and regularly-changing text, involving a considerable level of pre and post-editing.

However, for the purposes of this research MT relates to the free translation programs that are available such as the Google translator (Google Language Tools 2008). In some cases, a link is provided on a web site enabling the user to click and access the machine translation of the current page. It is also possible to enter the URL of the page required in translation into the Google 'Language Tools' program, decide which language it should be translated into, and click to translate the page in question. This method of translating web sites is possibly sufficient to give the user the 'gist' of the ST. However, it is unsuitable, for example, for the translation of a company web site that is intended to professionally showcase the company. In addition, as the translation is automated, the issue of cohesion cannot be addressed.

2.6.2 Web site localization

'Localization' is described by the Localisation Industry Standards Association (LISA) as the process whereby a product is 'linguistically and culturally appropriate to the target market where it will be used and sold' (LISA/AIIM 2001), the 'product' being traditionally a software program. The business of web localization emerged from

software localization and refers to the recreation of web sites in particular language versions (O'Hagan and Ashworth 2002 pix). Similar to software localization, web localization is a technically-driven process that aims at standardization in terms of style and terminology, and where language quality is 60-70% dependent on terminology (Scarpa 2002). The key milestones in a web localization project include engineering, enabling, translation, testing and publishing (TILP 2005). O'Hagan (2003) defines localization:

One way to define localization may be to see how it differs from conventional translation. For me, that difference lies in the nature of the content with which it deals. Localization can be seen as an industrial process applied to content that is predominantly in digital form and needs to be adapted to target market requirements.

Cheng refers to the process of providing multilingual content on a web site as 'Globalization', a process that has three main areas: translation, workflow and delivery (2002 p34). These areas cover the three overall objectives in the provision of a multilingual site of front-end (involving 'linguistic' and 'cultural/marketing' issues); back-end (technical and managerial issues) and long-term (future web site needs and changing functionality) (2002 p30-32).

Apart from issues of terminological equivalence and the adaptation of text boxes and time and date conventions, other translation strategies involving issues such as context and cohesion do not appear to be a fundamental requirement of the translation brief with this approach. It is, therefore, questionable to what extent such issues can be dealt with in this technically-driven environment. Pym expresses his concern that, in the

final analysis, linguistic issues are no longer decided by the translator but by the marketing expert or the engineer:

The sheer size of most one-to-many communication projects means there is an increase in the hierarchical control and standardization of translation. 'Localization' may superficially mean 'translation plus adaptation', but these two aspects are increasingly separated. The various translation-memory programs and localization tools restrict the translator's decisions, returning strict translation to the paradigm of phrase-level equivalence, and leaving adaptation to specialists in marketing and engineering (2006 p750).

Cheng suggests that 'enabling a sophisticated corporate or e-commerce Web site to be multilingual and multiregional is a complex and subtle undertaking' (2002 p29). In addition, given the cost involved it can reasonably be assumed that such a project is financially out of reach of many SMEs. Consequently, companies in the SME sector are often more likely to employ the third or fourth approach (see 2.5.3 and 2.5.4 below) when providing multilingual content on their web sites.

2.6.3 Translation company

In this case, the text for translation is provided to the translation company which sends it to the translator or translators, in the case of a web site that requires translation into a number of languages simultaneously. In the vast majority of cases, the translators used by a translation company are freelance (Boardman, DCU-LS, Minogue all 2008). In most cases, the translator receives the document in Word format and sometimes XML, and will return it in the same format to the translation company which forwards it to the client for uploading by the IT people. Ideally, the text will be reviewed and revised once it is in place on the web site and before going live. In reality, as mentioned in

Chapter I, the translation company often has no influence over the TT once it has been passed to the client.

2.6.4 Freelance translator commissioned directly

The fourth possibility involves, for example, a translator who may have carried out translation work for a specific client for a number of years and is asked to translate the web site which may have been set up recently or now requires content in a specific language. The translator will generally be provided with the ST in Word format and will return it in the same format. Where a source web site already exists, it can be used as a reference point for the translation. The translator may or may not be in position to suggest revisions once the text has been put on the web site.

Depending on which approach to the translation of a web site the translator is involved in, s/he may be familiar with all or some of a number of different programs, file formats and platforms (e.g. Word, PowerPoint, HTML, XML, TagEditor, Translation Memory), in addition to graphic and audio file structures. As outlined in Chapter I, this study is concerned with translation as a product as it is assumed, that regardless of the manner in which and the software with which web content is translated, there should be little or no difference in terms of the cohesion of the TT. The changing role of the freelance translator, s/he being the most likely person to carry out web translation, is also of concern in this project.

Much has been written about the translation of web sites as part of web localization projects (Pym 2004; Dunne 2006; McDonough 2006; Sandrini 2007; Pierini 2007).

Very often the focus in this case is on issues such as the ‘Globalization’ or the

'Internationalization' of web sites. Sandrini recommends that translation studies and the localization industry come together and pool resources as web localization is the perfect example for a functional approach to translation that is closely linked to economics and business (2007). McDonough (2006 p85) suggests that the process of localization adversely affects perceptions of Self and Otherness and that in some cases it may not be necessary to localize. Pierini (2007 p98) writing on quality in web translation, suggests that the general linguistic criteria of 'adequacy' and 'acceptability' should be integrated with genre and medium-related criteria (effectiveness) and medium (legibility) respectively, the overall objective of the exercise being 'to produce a "natural" TT without errors'.

The linguistic decisions taken by the translator no longer involve finding the most appropriate TL terminology for the given context, but, to return to the concept of Information Ecology posited by Rosenfeld and Morville, involves a full understanding of the three items in that ecology: 'user', 'context' and 'content'. The translator will, in essence, know and understand the language of the potential 'user' of the site, in the particular 'environment' in which the company or business functions and will be familiar with how the 'documents, applications, services and metadata that people need to use or find (a) site' work together to create a cohesive web site and ensure a good user experience in the TL (Rosenfeld and Morville 2002 p25).

2.7 Translation of the corpus web sites

A brief survey was carried out by the researcher in order to establish which procedure was used for the translation of the corpus web sites, for updating the TL content, the level of involvement of the web site owner and whether the TL content was tested before going live. The people responsible for each of the corpus web sites were contacted. In some cases this was a company commissioned to provide the web site presence of the city in question. In Munich this is the Portal München, Betriebs-GmbH & Co. KG. With other cities (e.g. Frankfurt, Garmisch-Partenkirchen) the city or town council is responsible for the web presence. From the ten requests for information sent to those responsible, replies were received from about seven of the corpus sites (Cologne, Frankfurt, Garmisch-Partenkirchen, Hamburg, Leipzig, Munich and Regensburg).

Alexander Esser for Redaktion Koeln-de stated that a translation company was commissioned to do the initial translation but updates are now carried out by employees at 'Redaktion koeln.de'.

Melanie Hartlaub, responsible for ICT at the City of Frankfurt, states that the initial translation for the Frankfurt site was carried out by a translation company but the updates are dealt with internally.

In the case of the Garmisch-Partenkirchen site, Ms Rappenglück, responsible for E-Marketing, replied as follows:

den derzeitigen Webauftritt hat eine freiberufliche Dolmetscherin übersetzt, im Bereich Pauschalen war auch mal eine Kollegin intern tätig, die Pressemeldungen übersetzt uns ebenfalls eine freiberufliche Übersetzerin

Ms Rappenglück also mentions that her office decides on the texts to be translated and uploads them to the web site without further consultation with the translator of the texts.

Torralf Köhler, head of marketing and sales at hamburg.de GmbH & Co. KG, reports that a combination of translation company, freelance translator and company employees were responsible for the initial translation of the web content. Both a translation company and freelance translators commissioned separately are responsible for the translation of updates. The updated content is sometimes tested before going live.

In the case of the Leipzig site, Elizabeth Hoppe reports that the initial translation of the web site was provided by a translation company and she reports the following in terms of updates to the TL text:

Für Updates sind die entsprechenden Fachämter verantwortlich, die die Inhalte in der deutschen Version bereitgestellt haben. Sofern es sich um geringfügige Updates (hauptsächlich Zahlen oder Veranstaltungen) handelt, werden die Korrekturen von den Mitarbeitern vorgenommen. Hier gibt es auch Hilfe zwischen den Ämtern untereinander, da nicht in allen Ämtern Übersetzungspotential vorhanden ist. Wenn größere Übersetzungen erforderlich sind, werden wieder Übersetzungsbüros eingeschaltet.

With reference to testing the TT prior to going live, Elizabeth Hoppe provides the following information.

Bei der Auswahl der Übersetzungsbüros war Bedingung, dass ein Muttersprachler eingesetzt wird und ein zweiter Übersetzer Korrekturlesungen vornimmt. Die Freischaltung der übersetzten Seiten erfolgt erst nach Prüfung durch die Fachämter.

The Munich site is the second largest in the corpus in terms of total number of pages which includes content in the SL and all TLs. Munich has approximately 560,000 pages and Berlin 1.1 million. Claudio Franchi, responsible for the Munich site, states that the initial translation was carried out by a freelance translator and all updates are also carried out by a freelance translator. In addition, updates are tested before going live.

Elizabeth Knott, Press Officer for the City of Regensburg, states that, with the exception of the tourist information pages on the Regensburg web site, a local freelance translator provides the translated text, which is checked by the web editor before uploading.

It is reasonable to assume that all or at least the great majority of translation companies in Germany, as in Ireland, use freelance translators as opposed to employing staff translators themselves. As seen above, therefore, freelance translators are involved in the translation of each of the seven corpus web sites that responded to the survey. This is a strong indication of the level of involvement and the importance of the freelance translator in the translation of web sites in this sector.

A second point of interest that emerged from the above information received about the corpus web sites is the involvement of internal employees in updating the TL content. While the level of involvement is not evident, this approach suggests that the texts in question are translated in isolation and are possibly not considered in terms of the overall cohesion of the web site in question.

2.8 Conclusion

Chapter II provided an outline of traditional approaches to cohesion. The nature of the web site was examined, and specifically the instances of text on a web site. The concept of 'hypertext' and its consequences, among other things, for reading patterns on a web site are examined and three areas of user interaction with a web site that set it apart from reader engagement with conventional text are identified. Aspects of traditional approaches to cohesion are found to be relevant to instances of continuous text on a web site. However, they are insufficient to account for cohesion in other instances of text. It is shown that aspects of visual composition are vital for the cohesion of a web site. A definition of web site cohesion is proposed. Some approaches to the provision of multilingual content on a web site are discussed. The results of a small survey carried out among the corpus web site owners/operators with regard to the translation of the corpus web sites are presented.

Chapter III presents the model for the analysis of web site cohesion with specific emphasis on aspects and devices that are relevant to the translated web site. The model draws on the three areas of user interaction that were identified in Chapter II: Online Search, Navigation and Page Content. A general outline of each of the three areas is provided, a more detailed discussion of the cohesive devices in those areas and details of how the model is tested on the corpus web sites.

Chapter III

The Model for the Analysis of Web Site Cohesion

Chapter III

The Model for the Analysis of Cohesion

3.0 Introduction

Chapter III presents a model for the analysis of cohesion in the translated web site. It is divided into three sections each of which deals with one of the three key areas that were identified in Chapter II as being central to web site cohesion from a translation perspective: *Online Search, Navigation and Page Content*. Each of the three sections consists of a general introduction to the respective area, a discussion of the individual cohesive devices and an outline of the application of the model to the research corpus.

3.1 Background

As was established in Chapter II, hypertext has led to the emergence of three new areas of user interaction with the web site as text that are referred to for the purpose of this study as ‘Online Search’, ‘Navigation’ and ‘Page Content’. It was also established that while traditional approaches to cohesion could be applied to instances of continuous text on a web site, this is not entirely the case with other instances of text on a web site. The model, therefore, draws together and builds on aspects of these approaches to cohesion, for example ‘exact repetition’ (Halliday and Hasan 1976), ‘simple repetition’ (Hoey 1991) and ‘repetition’ and ‘collocation’ (Halliday and Matthiessen 2004). The concept of ‘repetition’ in Martin and Rose is

appropriate only in the context of keyword metatags, where terms are included in the metatag (e.g. visit/visiting) in an attempt to provide the web spider with variations of a given keyword.

Cohesion on a web site is, in essence, lexical. However, lexis alone cannot account for cohesion of the textual elements under investigation given the spatial and visual aspects of a web site and its multi-linearity. The model, therefore, draws firstly on the three elements that make up the compositional metafunction in visual communication; framing, information value and salience. It also draws on aspects of Nielsen's guidelines on 'Usability', Callan's approach to 'Search Engine Optimization' and guidelines on 'Information Architecture' by Rosenfeld and Morville. Together, they create a model for the analysis of cohesion in the translated web site.

The first section in this model addresses the topic of search using the Google search engine and is referred to as 'Online Search'. It outlines the nature of this form of access to Web content and details the criteria applied by the search engine, Google when deciding which web sites to display in a list of search results. Against this background, and drawing on Callan's guidelines on Search Engine Optimization (2005), the devices that are capable of functioning cohesively in the translated text are described and the process of analysis detailed.

The second section of the model deals with 'Navigation' and points up the importance of navigation systems in a web site medium and their significance for the general cohesion of the translated web site. The aspects of navigation that are potentially cohesive are outlined and the process of application of this part of the

model to the corpus is detailed. This section draws heavily on Rosenfeld and Morville's framework for the organisation of web sites as outlined in their 'Information Architecture for the World Wide Web' (2002).

The final section of the model for the analysis of the cohesion in the translated web site deals with on-page user behaviour and is referred to for the purpose of this study as 'Page Content'. It deals with the potential for cohesion of items such as the teaser text. This section draws on recommendations in the context of website usability that contribute to web site cohesion.

3.2 Online Search - an overview

There are three main ways of accessing content on a web site:

- by inputting a specific web site address or URL into the address line of an Internet browser such as Internet Explorer
- by clicking and following a link from one web site to another
- through online search.

Online search, which is a keyword-based search for information on the Web, is the most popular means of accessing web content, with 83% of Irish adult web users accessing web sites using this method (Ámarach Consulting 2008). A survey carried out by Internet media and market researchers Nielsen//NetRatings (2006a) shows an increase of 55% in the number of online searches carried out in the US in the month of December 2005 compared with December 2004, while in the same period the number of people connecting to the Internet increased by a mere 3%.

In online search, the user enters a term or terms, also referred to as a search query, into the search box of a search engine such as Google. A search engine is a computer software program designed to help the user to find information on the World Wide Web with the help of keywords (USDHHS Glossary 2006a). In the context of online search such search terms are referred to as keywords, a keyword being a word 'that is used as a reference point for finding other words or information using a search capability in a web site' (ibid.). A keyword phrase consists of multiple keywords grouped together that are likely to co-occur in a search query and that have a greater collective strength than individual keywords (Emarketing 2006). In reply to a search, the search engine returns a list of active links, referred to as 'search results', to web pages with content that corresponds to the search query. The search results returned by the search engine are linear, prioritized and are grouped in pages of ten results (Nielsen and Loranger 2006 p140) referred to as a SERP. The links on the SERP are listed according to the importance of the pages to which they are connected. The level of importance of each page is established by the search engine itself. Based on the information about each destination web page provided to the user in the individual search result, the user decides whether to access one of the web pages listed or not.

Cohesion in Online Search occurs ultimately where the user is facilitated in accessing relevant and valuable content, in line with his/her requirements as expressed in a search query. It is brought about by the matching by a search engine of the keywords of a search query with keywords provided in the body text of the

destination web page and in certain HTML tags defined for connection with that page.

In web site translation the aim is to ensure that an adequate level of cohesion can be achieved in Online Search. This involves treating the TT as a text in its own right and not merely as an extension of the ST that needs no optimization techniques of its own. It involves applying optimization techniques, even though they may not be present or sufficient in the ST. Decisions as to the keywords and other terminology selected for the TL content is not solely a matter of choosing the most appropriate terms in the given context. It involves researching and establishing user language in the context. This refers to the keywords and other terms commonly used by the Web users when searching for specific content in a specific context. The terminology of the user is used and placed strategically throughout the translated text.

Online search is considered by Nielsen to be a 'fundamental component of the web user experience' (2005). According to Pasternack (2005), search behaviour has changed over time. In the initial years of the Web, search engines were used in a more random and unstructured manner, more 'browsing' than 'searching'. Today, the web user uses a search engine not only to search for new content but very often simply to re-visit familiar web sites and content. 'Search as a utility has become deeply ingrained into people's everyday lives' (Nielsen//NetRatings 2006b). Nielsen and Loranger suggest that users no longer search for sites which they then explore in depth. Users are now looking for specific answers, where 'search engines' have become 'answer engines' (2006 p36). The results of a 2006 survey into search engine user behaviour among the online population in the USA emphasise the

importance of adequately addressing the issue of online search. Some interesting changes in behaviour over the years in comparison with similar studies carried out in 2002 and 2004 are also revealed (iProspect):

- 62% of search engine users click on a search result within the first page (48% in 2002) and 90% click a result within the first three pages (19% in 2002).

This underlines the importance of a web page achieving a high ranking with search engines, which will influence the position of that web page in the SERP and the likelihood of users seeing it listed and having the opportunity to access it.

- 82% of searchers will re-launch a search with the same search engine but will add more keywords to subsequent searches – suggesting that the web writer/translator should target longer, more specific keyword phrases as opposed to individual keywords.

- 36% of search engine users believe that the organizations or entities that appear at the beginning of a search results list are the top organizations in that field. 39% have no opinion on this and 25% do not believe this. The company, iProspect, that carried out the survey suggests that there is still a considerable proportion of search engine users who believe that ‘industry leadership’ is one of the criteria used by the search engine when deciding which web sites to display at the top of the list.

The demand today for direct and instant access to relevant content is compounded by the exponential growth of the Web over the past few years. As of November 2006, there were over 100 million web sites registered, double the 2004 figure of 50

million (Netcraft 2006). This situation has had serious implications for web sites in terms of online search as web site owners are competing for the attention of potential web users with increasing numbers of web sites. Against this backdrop, it is crucial that the issue of cohesion in online search be addressed in the translated web site as, on the one hand it will contribute to the ranking of web pages with search engines and more importantly will provide the user with a clear indication, through the search result, of the exact content of the web page to which the result links.

Otherwise, despite the fact that the SL content may enjoy a very high ranking, and therefore, high level of visibility with Google, it could happen that potential TL users of the web site will remain oblivious to what is probably relevant and valuable content. Search engine developers are constantly forced to revise and improve the methods they use to establish the importance of one web page over another, a factor that will decide how high up on a SERP a web page will be listed. However it is not only the search engine developers who can influence how a page is ranked by a search engine. The web writer, and consequently, the web site translator have an important role to play in adapting relevant aspects of web pages to correspond with the requirements of search engines, while at the same time ensuring that the needs of the web user are met.

3.2.1 The Google search engine

The decision as to which web pages Google will list in a SERP and the order in which these pages are listed is based on two criteria. The first is referred to as ‘page ranking’, which refers to the overall importance of a web page from the point of

view of Google. The second involves the relevance of the content of a given web page in relation to the term or terms used by the user in a search query.

The software program, PageRanktm is at the heart of Google search. PageRanktm uses the link structure of the Web to establish the relative value of individual web pages. It involves evaluating incoming links to a page. A link from page *A* to page *B* is considered to be a vote by page *A* for page *B*. This is based on the premise that a web page will generally only include a link to another web page that is relevant and trustworthy. A link from a page that is already considered to be important by Google is of greater value than a link from a less important page. Google gives a web page a value out of 10. A Google page ranking of 5/10 and upwards is considered to be a good ranking. The ranking of a web page is reflected in the actual position of that web page when it appears in a SERP (Whitespace 2006). The higher the ranking awarded to a page by Google, the higher that page will appear in the Google SERP, provided the content corresponds to the search query.

Google, however, does not rely solely on the number and value of incoming links to a page in order to determine the value of that page. It also analyses the content of a page and the content of pages linking to that page. It seeks to match the search query with identical terms and phrases that are found both in the page content and in HTML tags that have been defined for a page. Factors that play an important role include the location of keywords on a web page and in the HTML tags, the proximity of the keywords to each other and the order of these keywords in comparison with the order of the keywords in the search query (Vaughns 2007).

The information displayed in a Google search result is drawn from data that has been gathered by thousands of ‘computers’ that ‘crawl’ thousands of web pages per second. These computers are referred to as ‘web spiders’ or ‘web crawlers’. They read a web page from left to right and from top to bottom. Information gathered from web sites for Google is taken not only from the page content of a web page and the pages that are linked to it, but also from certain tags contained in the ‘HTML source document’ of the page that is to be listed in the SERP. The information gathered during a crawl is sent to the search engines for indexing. The web spider also follows links on a web page, both links between pages on the same site and from one web site to another site. Most importantly from a page ranking point of view, it records links that are incoming from pages on the same site and on other web sites and it also gathers information on the content of these linked pages. The importance of a web page, the size of the web site and the frequency with which a web page is updated will determine how often the page is subsequently crawled by the web spider (Web Design Ireland 2005).

3.2.2 Search result

A search result is, in essence, an overview of the content of the page to which it links. It is also an advertisement for that web page and an opportunity to attract the user’s attention ahead of the other competing web pages that are listed in the SERP. The search result in Figure 3.1 is number 1 in a list of 305,000,000 search results returned for the search query ‘Berlin’ on www.google.ie. It links to the English

homepage on the official web site for Berlin:



Figure 3.1: Search result with page title, description text and URL

(<http://www.google.ie/search?hl=en&q=berlin&meta=>)

The Google search result contains a number of different items of information, the two most important of which from a translator point of view are the first line of the result, which carries the page title, and the second chunk of text which is a description of the page content to which the search result links. The page title has enormous potential as a cohesive device as it firstly creates a strong link between the user search query, the search result and the destination web page. It also attracts the attention of the user as it is the most salient piece of text in the search result due to its greater font size and its colour in comparison with the other elements of text and to the fact that it is the first piece of information on the destination content to be seen by the user. The page titles in a SERP are in strong competition with each other as the user moves from one heading to the next, making a split-decision as to whether to access a particular web page or not. It is also the most important place for the positioning of keywords from a search engine point of view (Callan 2005 p13). Where the page title has not been defined for a web page or where the site has not yet been indexed by Google, this first line in the search result will contain the URL

of that page. The first line of a search result is an active link, allowing the user to click and directly access the web page.

The second piece of text in the search result as seen above is a brief summary or description of the page content to which it links. Google has chosen to use the first 132 characters of the description content metatag that has been defined for this

English homepage, which is:

```
<meta name="description" content="Welcome to Berlin's official internet site. Here you can find information about different aspects from Berlin: Hotels and Hotel Booking, Visitors Guide of Berlin, Entertainment in Berlin and Tickets, Political System and Local Authorities, Berlin Capital City, Business in Berlin" />
```

Figure 3.2: Description content metatag, English homepage, Berlin web site
(<http://www.berlin.de/english/>)

The description text normally expands on the information provided in the page title on the first line of the search result.

The final line of the above search result begins with the destination page URL, ‘www.berlin.de/english/’, which is not an active link. This is followed by the information ‘33k’, which refers to the size of the text portion of the page to which the search result links and gives the user an idea of download time (Google Search Results Page 2007). Finally, the ‘Cached’ link leads the user to the last ‘snapshot’ that was taken of the page by the web spider. The current page may differ to the cached photo if the page has been updated since the last time the page was crawled.

Figure 3.3 is the ‘cached’ photo of the destination page for the search result in Fig. 3.1:



Figure 3.3: Cached snapshot of destination page from search result in Figure 3.1.

(<http://66.102.9.104/search?q=cache:US0VTrhAARYJ:www.berlin.de/english/+Berlin&hl=en&ct=clnk&cd=1&gl=ie>)

The keyword or keywords used in a search query are highlighted in the cached snapshot, as can be seen with the word ‘Berlin’ which is the search query for the above example. This gives one a good indication of where the keywords are positioned on the page and which instances of keywords are ‘read’ by the Web spider and which are not read. The web spider is unable to read text on a web page that is defined as a graphic element. If the link in the search result labelled ‘Similar pages’ is clicked, Google will scout the Web for more information on the same subject.

3.2.3 Search Engine Optimization

The framework for the analysis of cohesion in online search draws on aspects of Search Engine Optimization (SEO). SEO refers to

the process of developing a website's structure and content to enable it to achieve the best possible rankings within the results being displayed on a search engine after a search is conducted for a particular term (Websearchworkshop 2007).

SEO is concerned with the prominence of web sites in organic search results.

Organic search is a search where results are based on the natural indexing of web sites as opposed to indexing based on paid advertising (Webopedia 2007c).

Nielsen and Loranger (2006 pp165-169) consider SEO to consist of three major categories, namely Linguistic SEO, Architectural SEO and Reputation SEO.

Linguistic SEO refers to writing for the web that speaks the user's language.

Architectural SEO has two objectives: ensuring that the pages of a web site can be indexed and having an appropriate link structure which will guide the web spider to the page content. Finally, Reputation SEO concerns the reputation of one's site from the point of view of the search engine. If a site offers a high-quality product, or interesting content or helpful services, other web sites will want to link to it, thus enhancing the reputation of that site.

Callan (2005 p50) considers SEO to involve two elements, 'off-page optimization' and 'on-page optimization'. Off-page optimization refers to

all of the ranking factors that are not located on (a) webpage, that the search engines look at when ranking a website (ibid. p23).

This includes the value of the web sites linking to a page, the number of linking web sites and the total number of incoming links to the web page. The incoming links to a web page can be from pages on the same site or from pages on a different web site (ibid. pp72-74). 'On-page optimization', on the other hand, involves the treatment of web site content in terms of keywords, keyword research, keyword placement and the creation of HTML tags (ibid. p13).

This section of the project is less concerned with those aspects of page ranking that require high-level incoming links from an external web site in order that the page will receive a high ranking. However, the issue of keyword links, which from the point of view of the destination page are inward links, is addressed in this section as opposed to in the navigation section. This section of the study is concerned with those elements over which the translator has a level of influence and which directly affect cohesion in the translated web site. This involves firstly defining high-quality keywords in the TL, placing these keywords strategically on the web page, creating keyword links, where appropriate, and finally optimizing the relevant HTML tags, all of which are vital for overall cohesion and a good user experience, enabling the user to access relevant content quickly and easily.

3.3 Online search – the model

The model for the analysis of web site cohesion in online search contains the following textual features of a web site that play an integral part in the creation of cohesion in the translated site:

- Keywords
- HTML tags
 - title tag
 - h1 header tag
 - description content metatag
 - keywords metatag

3.3.1 Keywords

In a web site context, keywords are the most important terms identified that describe the content of a web page. Ideally, keywords are researched and defined for all pages of a web site given that no two pages will have identical content. There will inevitably be some overlap as, for example, one keyword such as the product name could be relevant to all pages of the web site. Keywords are included in the body text, as navigational labels on the web page and also in the relevant HTML tags such as the title tag and the description content metatags. While it is not an absolute requirement on a web site, be it in SL content or TL content, to define and strategically place keywords on a web site, the failure to do so will have serious implications not only for cohesion in online search but also in navigation and page content.

Cohesion involving keywords in online search occurs where a link is created between the term or terms used in a search query and a web site that is returned in a SERP and that has content that corresponds to that required by the user. The link comes about through the repetition of keywords, that is to say in the search query and on the web page.

There is no requirement to define keywords for any page of a web site and given the increasing number of pages now available on even the smallest web site, it is often the case that keywords are not defined for all pages. However, they should be carefully researched and defined for at least the most important pages of a web site, that is the homepage and the main section pages. Regardless of whether keywords have been defined for the ST or not, they should be defined for the TT and placed in

the positions of maximum effect in order to contribute to the overall ranking of the web pages and to facilitate cohesion in terms of the user. Defining and placing keywords and the keywords metatag is addressed in the section entitled 'HTML Tags'.

Keywords defined for a web page reflect the user language, that is they are terms that are used or likely to be used by the Web user in a search query for content identical to that on the web site for which the keywords are defined. Nielsen (2006b) suggests that the value and visibility of many web sites are diminished by filling the tags and metatags with supposed keywords and terms that are identical to terms used on the web site but are not used by the user in an online search. Therefore, they will not contribute to the cohesion of the web site as a cohesive link cannot be created due to a lack of lexical correspondence. This issue is of particular significance for the translated web site. It is not sufficient to 'translate' by finding appropriate and equivalent terms in the TL as cohesion will only be aided if the terms chosen for the TT are ones that are likely to be typically used in the given context by the user, that is the TL audience.

Keywords are neither too general nor too technical. If a keyword is too general it may have to compete in online search with many other web sites that have the same main keyword, which will make it more difficult to achieve a successful ranking. On the other hand, if keywords are too technical or specific they may not be used as a search query by many people. This could result in a failure to create a cohesive link between the search query and the web site as the terms used in the query and on the web site do not correspond to each other. However, the level of specificity could

vary from one language to another and this must be taken into consideration when selecting TL terminology.

It is not uncommon for a company or organisation to use different names for one and the same product. This can occur, for example, where a product is given a specific name during its pre-development and development stages but is given a more suitable name when launched onto the market. On the company web site, the product is referred to using the in-house name, which however may not be known to the user carrying out a search using the market name. As such a term would be one of the main keywords for the company site, it must be chosen carefully and with the end user language in mind. Otherwise, there will be no cohesive link as the search engine will be unable to recognise the fact that one object is referred to using two different terms. Even if the search engine manages to return the result high up on the SERP, the information provided in the search result may still prove to be less than cohesive as the user may not recognize that the unfamiliar term refers to the same item as requested via a search query.

The use of promotional language which can be eye-catching and successful as part of an overall advertising campaign is less suited when defining keywords for a web page. Nielsen (2006b) suggests calling 'a spade a spade, not a digging implement. Certainly not an excavation solution', as the average user will generally choose the simplest way of describing a product. Nielsen recommends the use of 'old' words,

meaning words with which people are familiar, as

familiar words spring to mind unbidden. Thus, users are likely to employ old words when they boil down their problem to a search query, which is typically only 2-3 words long. (Nielsen 2006b).

The fact that a word is unexciting would indicate that it is frequently used (ibid.).

Once again, if the keywords defined for the page do not match those used in the search query, a cohesive link cannot be created as the search engine will fail to recognize that the two different terms actually have the same referent. It is tempting to coin new words or phrases, especially if such words or phrases are likely to become anchored in a language. If the use of new words is unavoidable, some familiar terms should be added in (ibid.).

Keywords are precise and unambiguous, which implies that their meaning should be instantly clear in the context within which they are used. Consistency of keywords is more useful to a search engine and to the user than a variety of synonyms or near-synonyms. In addition, the user is more likely to use the same terms when carrying out a repeat search for the same content.

Where web sites are adapted in some way for the TL audience it can happen that the key message on a page has shifted slightly in comparison to the initial message. TL keywords may need to be revised to reflect this change in emphasis. This also applies to content on a web page that has been updated. While the general message may remain the same, it may nonetheless be necessary to revise the existing keywords or even add in some additional keywords to those defined for that page.

The keywords that are defined for the translated web site are in the appropriate language, unless there is a sound reason for using a term in the SL or another

language. Keywords on a translated site that are in the SL or another language will not be recognized as keywords either by the search engine or the users and will be incapable of functioning cohesively.

Keywords are present on a web page in text rather than graphic format, as the web spider is unable to read non-text elements and will, as a result, fail to register these words as keywords (Google 2007d). SEO guidelines recommend placing the most important keywords as close to the left on a text line as is linguistically appropriate as this is the position of greatest impact from a user point of view. In addition, it is the position that is most advantageous from a search engine point of view (SEO-Tutorial 2007). Keyword density refers to the percentage of keywords on a web page compared with the total number of words on the page (Callan 2005 p17). If the keyword density is too high and/or if the keywords occur unnaturally, the page could be rejected by the search engine and will, therefore, not be listed in its SERP.

Keywords that occur naturally in close proximity to each other and that reflect the order in which they are present in a search query have increased weighting with search engines (Google 2007c).

Highlighting keywords on a web page, where appropriate, contributes to the creation of a cohesive text as salience promotes cohesion in an online environment by facilitating the user in scanning a page quickly and spotting such visual cues. A highlighted keyword also has a greater weighting with Google than a keyword that is not highlighted. However, the excessive use of highlighted keywords is counter-productive as it could be perceived to be keyword spamming.

3.3.1.1 Keyword links

A keyword link or hyperlink refers to a link label that consists in whole or part of a keyword or keywords (Dennis 2004). The main difference between standard keywords on a web page and keyword links is that standard keywords are terms describing the content of the current page while the keywords used as link labels refer to the content of the destination page. The cohesive power of the keyword hyperlink is underlined by Dennis (ibid.):

Hyperlinks can be a wonderful bridge between search engines and pages. The keywords included within a hyperlink are weighed more heavily than the same keyword hanging out in your copy text.

These links have a greater weighting than standard keywords because they set up an incoming link for the destination page. This adds to the importance of the page from a Google point of view and will most likely help to enhance the page ranking of that page. From the user point of view, a keyword link, where appropriately defined, creates a strong cohesive link for the user between a web page and the destination content of the link and therefore facilitates a good user experience in that the pre-supposition of the link text is realised.

3.3.2 HTML Tags

HTML or 'HyperText Markup Language' is a coded format language used for creating hypertext documents on the Web and for controlling how web pages appear (Gramene 2004). As mentioned in a previous chapter, each page on a web site has its own HTML code, also referred to as 'source code' which is contained in a text document. The HTML code contains instructions on the layout, font size, text and

background colour for that web page, among other things. It also contains tags and metatags. A tag is a piece of code that is embedded in a HTML document which indicates to the Web browser how text or images should be displayed (Webopedia 2007a). A metatag is a snippet of text that is used to provide information to search engine indexes about the content of the page for which it is defined (Emarketing 2006). The tags that are of particular relevance to the translator in terms of web site cohesion in online search are the ‘title tag’, the ‘h1 header tag’, the ‘description content metatag’ and the ‘keywords metatag’.

3.3.2.1 The title tag

The title tag in a source document contains the ‘page title’ which is a brief and precise summary of the content of the page for which it has been defined and which ideally contains only keywords that are relevant to the page for which the title is defined (Callan 2005 p50). The title tag for the German homepage of the Munich site reads:

```
<title>muenchen.de - Offizielles Stadtportal für München</title>
```

Figure 3.4: Title tag, German homepage, Munich web site
<http://www.muenchen.de/>

which contains the page title: ‘muenchen.de – Offizielles Stadtportal für München’.

The page title does not appear on the actual web page.

In the context of cohesion in online search, the page title has two functions. It is the first and, therefore, the most important source of information about a web page visited by the web spider (Searchengines.com 2006). It is understood by the spider to indicate the precise content of the page and the individual terms are considered to

be keywords. Secondly, the page title is the first line of text in a search result and is therefore the first indication to the user of the content of the page to which the search result links. Where adequately optimized, the page title is a strong cohesive device linking the information indexed by the search engine to a search query and ultimately to a relevant destination web page.

Search engines recognise up to 150 characters in a page title (Nielsen 2006b).

Generally speaking, Google appears to display about 80 characters in a search result.

The page title is a precise and meaningful indication of the page content for which it is defined. Users will even scan over search results that have achieved a high ranking in the SERP if the first line of the search result, that is the page title, is meaningless (Nielsen 2006b).

Some web sites give an indication in the page title of the position of that page in the overall hierarchy of the site. This is done by including a keyword from the individual sections or hierarchical levels. The Garmisch-Partenkirchen web site takes such an approach. Figure 3.5. shows the page title defined for the page 'Brauchtum & Tradition' which is the main page of the sub-section of that name and part of the main section 'Information & Service' in the overall web site labeled 'Garmisch-Partenkirchen':

```
<title>Garmisch-Partenkirchen – Information & Service – Brauchtum & Tradition</title>
```

Figure 3.5: Title tag of page 'Brauchtum & Tradition, Garmisch-Partenkirchen web site
(<http://www.garmisch-partenkirchen.de/de/8668d2c6-3cb2-340b-3a84-d4bea8d3cd7b.html>)

As each web page will normally have content that is for the most part unique to that page, no two pages on a web site have the same title. This is significant for cohesion

in a number of ways. The web spider may not be able to differentiate between the content of different pages where two or more pages on a site have identical page titles. Consequently, such pages may be incorrectly indexed by search engines which could result in the pages not being returned in reply to a relevant query in an online search. As a rule, the page title should contain only the main keywords for the page it represents. The fewer keywords that are placed in the page title, the more weight each individual keyword will have from the point of view of the search engine (Callan 2005 p13). As the user often only reads the first few words of search listings, the most important information or keywords are placed as far left in the text as possible (Nielsen and Loranger 2006 p276). Redundant or filler words such as ‘and’ or ‘which’, are not keywords, do not provide any additional information on the content of the page and are best avoided in a page title as they will be ignored by the search engine and are simply wasting valuable character space in the text (Callan 2005 p50).

In order to be able to function cohesively both from a search engine and a user point of view, a page title must be in the correct language. The search engine will be unable to match the keywords in search terms with those of the page title if the latter is in a different language. Secondly, even if a page with a SL page title is returned in a top position in a SERP, the user may well scan over this search result and move onto the next entry in the list as the page title may not be understood.

Where a page title has not been defined, the URL for the page will be displayed in its place: in a search result, in the reverse bar and in history and bookmark lists. The URL of a web page frequently consists of numbers and symbols and therefore

seldom provides useful information either for the web spider or the user as to the content of the page. Here again, the page title will be unable to function cohesively. The page title is the most valuable cohesive device in online search given that it is read and used as an important indicator of page content both by the search engine and the user. It must, therefore, be optimized in order to enable it to fully exploit its potential in web site cohesion. Page titles should be included and optimized for all pages of the TL content, regardless of whether they are defined for the SL pages. In addition, the SL page titles should not be translated blindly, that is to say they should not be taken at face value. It is possible that the content of the page for which they are defined is not adequately indicated in the SL title or that the keywords selected a) do not reflect user language or b) are not sufficiently accurate or specific. In order to ensure that the page titles in the TL are precise and indicative of content, the starting point for the provision of TL page titles should be the page content and the user language and not the SL page title.

In addition to its function in terms of online search, the page title has three other functions which could be treated in the section 'Navigation' but will be dealt with in this section. Firstly, the page title is displayed in the reverse bar of the Internet browser where it provides the user with information on the content of the current page and consequently with a sense of orientation and context. The following is the reverse bar for the homepage of the Munich source web site:



Figure 3.6: Reverse bar with page title: German homepage, Munich web site
(<http://www.muenchen.de/>)

The page title is also used as the reference label for web pages stored in the history list. A history list is a list of web sites and their individual pages that have been viewed by the user over a specified period of time (Expedite 2007). The pages are categorized according to domain and can be viewed according to date, the frequency with which they are viewed, alphabetically or the order visited on a given day.

Figure 3.7 is a list of pages accessed on the Leipzig site and saved to the history list:



Figure 3.7: History list: Leipzig web site

Unique page titles are important for history lists as it is not possible to edit the text in any way before it is saved. If unique page titles are not used, a user who has accessed a number of pages on a particular web site on the same day, and wishes to access one or more of these pages at a later date, will have to open each page in the history list to check the content as it will not be evident from the history list label or page title. A cohesive link is not possible between different pages with the same page title, and therefore, the same keywords in the title, and a number of different pages, as the page title does not correctly indicate the content of the page to which it links.

As is the case with the history list, the page title can be used as a reference when saving a page to the bookmark/favorites list. The bookmark/favourites list is a list of web pages and sites that have already been visited by the user and that are stored and labelled for easy repeat access (Expedite 2007). The page title that is saved to the bookmark/favourites list can be edited or entirely replaced by the user. A list of some of the web sites in the research corpus that have been saved to the bookmark/favorites list is seen in Figure 3.8:

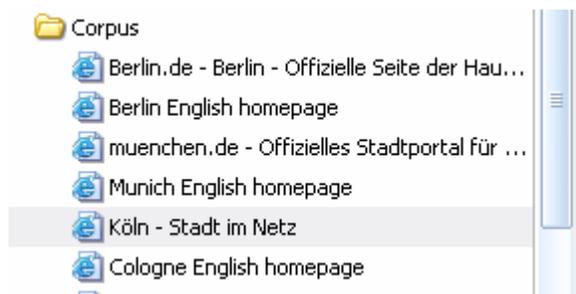


Figure 3.8: Bookmark/favorites list: corpus web sites

Unlike the history list, the page titles for the bookmark/favorites list can be edited, as is the case with the titles for the English homepages of the corpus web sites all of which have been saved using a standard title ‘City + English + homepage’, as seen above.

3.3.2.2 The h1 header tag

The text that is placed in the h1 header tag will appear on the visible web page for which it has been defined, unlike the page title. The h1 header tag reflects the main content of the page, as in the following tag defined for the German homepage of the

Cologne web site on 6 October 2007:

<h1>Die Top-Tipps zum Wochenende</h1>

Figure 3.9: h1 header tag: German homepage, Cologne web site
(<http://www.koeln.de/>)

This text appears as a heading on the web page as seen in Figure 3.10:



Figure 3.10: h1 header tag: 'Die Top-Tipps zum Wochenende', Cologne web site
(<http://www.koeln.de/>)

The h1 header tag plays a role in cohesion both from the point of view of the search engine and also of the user. Similar to the page title, it provides important information to the web spider about the content of the page for which it is defined. In online search a cohesive link comes about where the information contained in this tag is successfully matched with the page title and page content enabling the search engine to return a link in the SERP to a page with relevant content. In order to best serve the web spider, this tag should be positioned as far towards the top left-hand side of the page as possible (Callan 2005 p14). Cohesion is enhanced for the user

where the text of a search query corresponds to this text and is displayed prominently on the relevant web page when it is accessed.

Ideally, the h1 header tag will be a replica of the page title which is defined for the same page. The features apply as listed in section 3.3.2.1 above for the page title.

3.3.2.3 The description content metatag

This metatag is an objective and precise description of the page content and contains the main keywords for the page for which it is defined. The text of this metatag does not appear on the web page but can be included in a search result that carries a link to the page for which this tag is defined. The following example is the description content metatag defined for the German homepage on the Frankfurt web site at

www.frankfurt.de:

```
<meta name="DC.Description" content="Frankfurt: Die offizielle Homepage der Stadt Frankfurt am Main mit Informationen zum Leben in Frankfurt am Main, zu Tourismus, Kultur und Wirtschaft." />
```

Figure 3.11: Description content metatag: German homepage, Frankfurt web site
(http://www.frankfurt.de/sixcms/detail.php?id=stadtfrankfurt_eval01.c.317693.de&template=hp_flash)

Up to now, it was believed that the description content metatag should be no longer than 200 characters, this being the maximum length of text displayed by most search engines (Nielsen 1999 p233). However, it would appear that Google increasingly prefers to display a maximum of two lines of descriptive text which equates to approximately 140 characters. There are two principal uses for the description content metatag. It can be used by a search engine as the description text in a search result. It is also frequently used as a descriptive tag for web pages that are included in a web directory. The Yahoo Web Directory, for example, is a subject-based

directory that lists and provides links to web sites in many different categories (Yahoo Directory 2007).

Ideally, the text of this metatag indicates exactly what differentiates the content of the page for which it has been defined from that of other pages or websites in its category (DMOZ 2005). As the page description is read in conjunction with the page title, it is not simply a repetition of the page title but a true description of the page content. It contains the most important keywords for the page and is not simply a list of keywords. It contains no redundant text. The description content metatag should be in the language of the page content for which it is defined as otherwise it will fail in its role as a cohesive device as it will not be understood either by the user or the search engine.

A description content metatag functions as a cohesive device for the search engine by providing a lexical match with the page title and page content for the web spider. From a user point of view, it creates a lexical match with the search query and also provides additional information on the page for which it is defined.

Frequently, Google will not use a description content metatag in a search result. It is unclear whether the reason relates to the length of the text or the terminology used in the text. In such cases, Google creates its own description text for use in a search result. The description text in this case is created in one of two ways. Firstly, it can be a combination of snippets of text that contain the terms used in the search query,

as is the case in Figure 3.12. below with the search query ‘visit munich’:

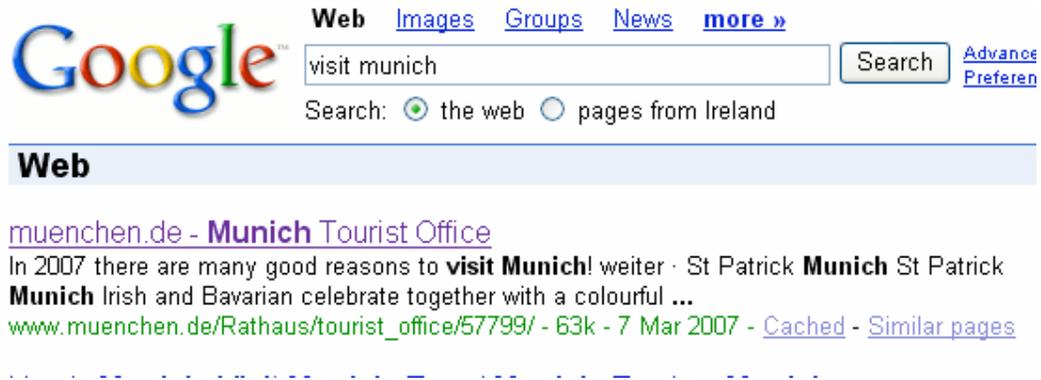


Figure 3.12: Description text: for search query ‘visit munich’
(<http://www.google.ie/search?hl=en&q=visit+munich+&meta=>)

These snippets can be drawn from any part of the page content, for example the body text, the navigation bar or even from the copyright notice at the bottom of the web page. They can also be drawn from web pages that are linked to the page in question. The result is at times less than coherent as the snippets are taken out of context. Consequently, they cannot provide an accurate description of the content of the page to which the search result links and therefore fail to function as a cohesive link to the web page in question.

In addition to drawing on pieces of text from the page in question to create a description, Google also draws on Web directories for the description text. In particular Google uses DMOZ, the Open Directory Project, for many of its search results (Web Workshop 2007). DMOZ is a comprehensive categorized directory of web pages and web sites. New pages and web sites are reviewed by volunteers before being added to the directory (ibid.). The ‘descriptions’ in such directories are

at times more precise and descriptive than the description content metatags that are defined for the page, and therefore more successful as a cohesive link when used in a list of search results.

While there is no guarantee that Google will use the description text that has been defined for a page, the likelihood is greater that it will be used if it is a true description of the content, is of sufficient length, contains the most important keywords reflecting the content of the page and, most importantly, contains keywords that are used in the search query and that, therefore, reflect the language of the user.

3.3.2.4 The keywords metatag

The keywords metatag is a list of keywords or keyword phrases that reflects the content of the page for which it has been created, as is seen in the examples in Figures 3.13 and 3.14.:

```
<meta name="keywords"
content="Stadtportal,Landesregierung,Senat,Bezirks&auml;mter,Beh&ouml;
rden,Verwaltung,Hotel Berlin,Ticket Berlin,Hotel in Berlin, Ferienwohnung
Berlin,Restaurant Berlin,Charlottenburg,Friedrichshain,
Hellersdorf,Hohensch&ouml;nhausen,K&ouml;penick,Kreuzberg,Lichtenber
g,Marzahn,Mitte,Neuk&ouml;lln,Pankow,Prenzlauer Berg,
Reinickendorf,Sch&ouml;neberg,Spandau,Steglitz,Tempelhof,Tiergarten,Tre
ptow,Wedding,Wei&szlig;ensee,Wilmersdorf,Zehlendorf,Touristeninformati
on" />
```

Figure 3.13: Keywords metatag: German homepage, Berlin web site
(<http://www.berlin.de/>)

```
<meta name="keywords" content="Landeshauptstadt, Muenchen,
Stadtportal">
```

Figure 3.14: Keywords metatag: German homepage, Munich web site
(<http://www.muenchen.de/>)

The purpose of the keywords metatag is to provide additional information to the Web spider about the page content for which this metatag is defined (SearchEngineWatch 2007). In the early days of online search, the value of a site was based on the number of times a keyword was present on a web page or in the HTML code for a web page. This led to the practice of ‘keyword spamming’ whereby some web writers filled their page content and the HTML tags with multiples of the most important keywords. Google, in particular, dislikes keyword spamming and penalises it by removing the web site in question from the Google listings (Google 2007d). The keywords metatag was consequently devalued in the eyes of many search engines (Ariadne 2007). It is generally thought, therefore, that the keywords metatag is no longer used by all search engines. Nonetheless, this metatag should still be completed as part of good design practice as it is still used by some search engines (Vaughns 2007).

It is not possible to say for certain whether keywords have been defined for a web page or not. Therefore, for the purposes of this study it is assumed that where a web site has defined a keywords metatag for individual pages, the keywords contained in that metatag are deemed by the web writer to be the most important terms relating to the page in question.

3.4 Online Search - applying the model

The first part of this analysis involves establishing the main keywords that are used in online searches in connection with the cities and towns in the corpus web sites. This is carried out with the help of ‘GoodKeywords’, which is freeware software

owned and operated by Softnik Technologies (GoodKeywords 2007). It is used for researching the most suitable keywords and keyword phrases for individual pages of a web site. For example, the search was carried out for all keywords used in connection with the word 'Cologne'. GoodKeywords then offers a list of the most common terms used in search results in conjunction with this term, as seen in Figure 3.15 below:



Figure 3.15: Result for keywords linked to the term 'Cologne', GoodKeywords.

One notable feature of keywords in an online environment is the pre-dominance of nouns, which is evidenced by the list of keywords related to the word 'cologne' displayed in Figure 3.15 above.

A second software program, Widexl, also a freeware program, is owned and operated by Widexl.com Internet Solutions (Widexl 2007). Widexl has, among other things, a keyword ranking analyzer and a metatag analyzer which draw on the search

engines Google, Yahoo and MSN for information on how tags and metatags and keywords are ranked. The keyword ranking analyzer enables the user to check the ranking, if any, of any keyword for a specific domain with a specific search engine. This is useful, for example, for checking the ranking of a web page for certain keywords and for monitoring the changing rank of the page for that keyword or keywords. Figure 3.16 below examines the keyword phrase ‘hamburg entertainment’ on Widexl:

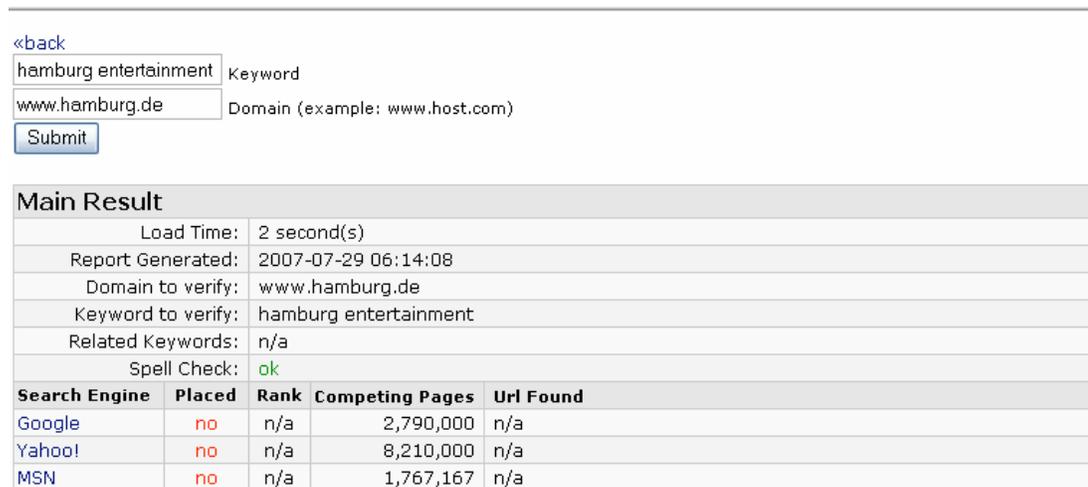


Figure 3.16: Widexl result of keyword analyzer for ‘hamburg entertainment’
 (http://www.widexl.com/seo/keyword-ranking.html)

The information in the columns labelled ‘Placed’ and ‘Rank’ indicates that the keywords phrase ‘hamburg entertainment’ has not been placed in the SERP or given a ranking by the search engines Google, Yahoo! and MSN. The keywords ‘munich entertainment’, on the other hand, have already been ranked by Google and MSN

and would be placed in fourth and eight place respectively in a SERP, as seen in Figure 3.17 below:



Figure 3.17: Widexl result for keyword analyzer for ‘munich entertainment’
(<http://www.widexl.com/seo/keyword-ranking.html>)

In effect, the Munich site would appear as number 4 in the SERP in reply to a search query with the terms ‘munich entertainment’.

The metatag analyzer analyzes the relevance of certain HTML tags and metatags defined for a specific page with reference to the content of that page. The score is

expressed in percentage of relevance, as seen in Figure 3.18:

Meta Tag Analyzer

URL
 Spider Robots (optional)
 User Agent (optional)

color = *Error* color = *Warn* color = *Good*

Status.		
PageRank:	<div style="width: 60%;"></div>	(6)
Status:	200 OK	
Web Server:	Microsoft-IIS/5.0	
Content Type:	text/html, text/html; charset=utf-8	
Charset:	UTF-8	
Content Length:	23349	

Meta tags report for: http://www.leipzig.de/		
meta tag	length	value
Title:	13	Stadt Leipzig
Keywords:	71	Stadt Leipzig,Stadtverwaltung,Behörden,Verwaltung,Touristinformationen

Meta tags analysis.	
Title:	<p>Title: <i>Title contains no errors.</i> This tag contains 13 characters.</p> <p><i>Title relevancy to page content is excellent.</i> The Title relevancy to page content is 100%.</p>
Description:	<p>Description: <i>Found no description meta tag.</i> The description meta tag is used by most of the search engines. Consider to add a description meta tag.</p>
Keywords:	<p>Keywords: <i>Keywords meta tag relevancy to page content is very poor.</i> The keywords meta tag relevancy to page content is 33%.</p>

Figure 3.18: Widexl result for metatag analyzer for 'http://www.leipzig.de'
(<http://www.seocentro.com/tools/search-engines/metatag-analyzer.html>)

As can be seen, the page title is considered to be very relevant and a description content metatag has not been defined for the page. In addition, the keywords listed are considered to be less than relevant. This is because the majority of the keywords listed are not included on the page for which the keywords list has been created.

Widexl assesses the relevance of metatags based on the level of repetition of the keywords that are placed in the keywords metatag and those placed in the page content. However, it does not take into consideration keywords that are typically used by the user when searching for information on the specific subject area. Therefore, a keyword that is considered to be very relevant and receives a high percentage may turn out to be quite irrelevant because it is never used in a search by the user. A cohesive link cannot be created between the user query and the web site in question. This in itself indicates the need to research user language before translating such tags as the keywords defined for a web page may not indicate user language.

Four keywords that are common to all web sites in the corpus are selected for use in an online search. The keyword suggestions made by GoodKeywords are based on the most common user search terms drawn from search engines such as Overture.de and Overture.uk, which are part of the Yahoo! network. The terms that were most common to all the cities in the corpus were 'hotel', 'restaurant', 'things to do' and 'entertainment'.

Five searches are then carried out for each city. The first search is a single-term search using the main keyword for each site, which is the city name. The following four searches on each site are multi-term searches that consist of the name of the city + one term each of 'hotel', 'restaurant', 'things to do' and 'entertainment'. Online searches for English content as part of this research project were carried out on www.google.ie and, where necessary for reasons of comparison, searches for German content were carried out using www.google.de.

On the basis of the SERP that is returned by Google, the ranking (where applicable) of the corpus web site is established. Where a web site is not returned in the first three pages of the SERP, Widexl is used to ascertain whether the web page has been ranked by Google at all. If it has, the ranking position is noted. Regardless of whether the page is listed in the first three pages of the SERP the individual elements of the search result are examined for keywords, keyword order, relevance to the destination content, TL, page title and description content metatag. The potential for cohesion of the lexis in the search result is ascertained.

3.5 Navigation – an overview

As discussed in Chapter II, navigation as a concept in verbal text takes on an entirely new meaning in a web site environment. It is an integral part of the functioning cohesion of a web site and is therefore vital for a good user experience.

Navigation tools of old have helped us to

chart our path, determine our position and find our way back. They provide a sense of context and comfort as we explore new places (Rosenfeld and Morville 2002 p106).

It is no different with the navigation systems on a web site, which act

as its road map, with clearly marked destinations, roads that intertwine, and even suggested routes (Timberlake 2000).

On a web site, the user has to contend with problems that are not there in the physical world: no sense of scale, no sense of direction, no sense of location (Krug 2000 p57). A comparative study by the Poynter Institute that looked at reading behaviour in print and online media found that in the print medium, large headlines

and photos received most attention. Online, however, readers tended to immediately look for navigation bars and teaser texts (2007). Unlike reading a conventional text which progresses in linear fashion from beginning to end, entering the world of hypertext can transport the user right to the heart of an unfamiliar web site. The hyperlink is undoubtedly one of the most significant cohesive devices on a web site as it allows the user to fully exploit the multi-linearity of a web site and to travel to any and all corners of the Web and in any direction or order decided by the user.

The Web's hypertextual capabilities removed any limitations on the direction in which the user can move (not just hierarchically, up/down, laterally etc.) allowing tremendous freedom of navigation (Rosenfeld and Morville 2002 p111).

Through links from remote web pages and search engine results, the user often completely bypasses the 'front door' or main page of a web site (ibid. p110). Users sometimes print off pages from a web site for use at a later stage. If the relevant contextual details such as the URL and/or an indication of the hierarchy of the web site are not available on the print out, the user may be unable to revisit that page or site. Rosenfeld and Morville claim that 'Context is King' in the design of navigation systems (ibid.). Good navigation systems 'give us a strong sense of location and smooth our path through the myriad pages of a web site, allowing us to move to another web site and just as easily to return to the original site' (ibid.). The consistency of features of visual communication such as colour, layout and graphic icons play an important role in the overall cohesion of a web site. However, above all text-based hypertextual navigational features such as navigation bars, hyperlinks and sitemaps are capable of functioning cohesively. They are fundamental to web

site cohesion in that they enable the user to move quickly and effortlessly through a web site and from one web site to another. These navigation tools are not cohesive *per se* but can become strong cohesive devices depending on the text label that is applied to them and their position in the overall visual-spatial composition.

3.6 Navigation – the model

Rosenfeld and Morville suggest that navigation on a web site generally consists of two types of navigation that can be further divided into ‘subsystems’. The first type of navigation is the category of ‘embedded navigation systems’ consisting of global, local and contextual systems which, are so to speak, ‘wrapped around’ the content of the site (2002 p107). They provide context and flexibility, enabling the user to know at all stages where s/he is on a site. The second type of navigation system is referred to as ‘supplemental’. The elements in this system provide an alternative means for the user to access areas from within a web site. Examples of elements of a supplementary navigational system are the sitemap, site index and web site-specific search feature.

3.6.1 Embedded navigation systems

3.6.1.1 Embedded navigation systems - global

As indicated in the name, a global navigation system is intended for inclusion on all pages of a web site. The most common form of global navigation system is the *navigation bar*, which is a bar placed either horizontally across the top of a web page or vertically down the left side of the web page or, less frequently, down the right-hand side of the page. The navigation bar contains individual links called ‘tabs’ that

when clicked, link to the key areas or sections of the web site. The links can be textual or graphic icons or a combination of both.

The navigation bar is a particularly important cohesive device in that, as a global navigation tool, it is placed on all pages of a web site, firstly providing a sense of location and context and secondly, enabling access from any page to both the homepage and to all the main sections of the site. This is all the more important given that the user can enter a web site even on the most obscure page, without passing through either the homepage or one of the main section pages. The shape, size and colour of the global navigation bar vary from web site to web site. Some are highlighted through the use of framing, colour, size or style and stand out clearly on the web page compared with other textual elements. Other navigation bars are less striking and resemble hyperlinks rather than the global navigation tool that they are. Figure 3.19 shows the navigation bar on the German homepage of the Munich web site:



Figure 3.19: Horizontal navigation bar with vertical navigation tabs, Munich site
(<http://www.muenchen.de>)

In the above case, the main navigation bar is placed horizontally across the top of the web page. It is repeated in a different format down the left-hand side of the page.

While the text and the individual colour coding of the individual navigation tabs are repeated on the left, the tabs resemble normal hyperlinks as opposed to the main section headings for the site.

A *breadcrumb trail* consists of a row of links, beginning on the left with the label for the home page and continuing towards the right with each label reflecting a level in the hierarchy of that section. The breadcrumb trail indicates the current position of the user in terms of the overall structure of the site. It is considered to be an element of the global navigation system of a web site, although it contains labels that are part of the local navigation system. The breadcrumb trail in Figure 3.20 (in blue beneath the ‘Tourismus’ navigation tab) is an example from the Frankfurt web site:



Figure 3.20: Breadcrumb trail: Startseite > Kultur > Museen > Alle Museen, Frankfurt site (<http://www.frankfurt.de/sixcms/detail.php?id=3796>)

In the above example the link *Startseite* is the homepage, *Kultur* is a link to the main section of that name, *Museen* is a sub-section of *Kultur*, and *Alle Museen* is in turn a sub-section of *Museen*. With the exception of the link for the current page, all the others are active links enabling the user to return directly to the homepage or to any

other of the sub-sections listed. As an element of global navigation, the breadcrumb trail is an indicator of the site hierarchy and does not, therefore, reflect the actual path taken by the user on that web site in arriving at the current page. Given the multi-linearity of the online reading path, the user may have passed through many different sections on the site.

Krug (2000 p63) suggests that, in addition to a section indicator such as a navigation bar, a global navigation system should include a *home button* and a *site ID*. The German homepage on the Hamburg site has two home buttons both labelled *Startseite*, as seen in Figure 3.21:



Figure 3.21: Home button ‘Startseite’, German homepage, Hamburg web site
(<http://www.hamburg.de/>)

The example on the left hand side, beneath the word ‘international’, is accompanied by an icon of a house representing the ‘homepage’. Both ‘Startseite’ and the homepage icon are clickable. The second instance of *Startseite*, situated in the white central area above, appears to be a clickable cue but is not.

Many web sites today include a site ID, available on all or most pages. This is frequently the company logo or the name of the organisation. In general such a site ID is clickable, linking to the home page. In the above example, the site ID is the

text 'hamburg.de' together with the red icon to the right of the text. Both are clickable and link to the homepage of the site.

Indicators of foreign-language content on a web site are normally present as part of the global navigation system of the site. Consequently, the user will be aware of and able to access content in a language other than the current one, regardless of the page on which s/he has accessed the web site.

The two primary means of indicating the presence of foreign-language content on web sites are the icon and text. The most popular icon used is the flag, for example the US Stars and Stripes indicates the presence of content in the English language. Textual reference to foreign-language content is sometimes in the language itself, for example *Deutsch* or *Русский* , and sometimes in the SL of the web site. When a language button is clicked, the page generally defaults to the home or main page of the foreign language content.

3.6.1.2 Embedded navigation systems - local

A local navigation system serves to provide context for and links to web content in the immediate area. The local navigation system may be but is not always similar on each page of a web site. In any case, the link text will differ, reflecting the destination content to which it links. A local navigation system can consist of a list of links or individual links.

Figure 3.22 below shows an example of a *local navigation bar* for the ‘Tourismus’ section on the Cologne site:



Figure 3.22: Local navigation bar for ‘Tourismus’ section, Cologne site
(<http://www.koeln.de/tourismus/>)

A *menu bar* looks similar to a navigation bar, however, each tab of a menu bar offers a list of options which are clickable and each of which leads to related content (Webopedia 2006b). There are different types of menu bar, for example a *drop-down menu* or a *sequential menu*. A drop-down menu displays the various sub-sections of that section either when the section link is moused over or permanently once the user has entered that section. A sequential menu is one that displays

‘...choices from multiple levels in the menu hierarchy’ (USDHHS 2006b), as in

Figure 3.23:



Figure 3.23: Sequential menu, Garmisch-Partenkirchen site
 (<http://www.garmisch-partenkirchen.de/de/db41673c-3301-78c8-8a27-52cc9b4d45f6.html>)

This example illustrates the hierarchy of the web site beginning with the section *Freizeit/Sport*, its sub-section *Sehenswürdigkeiten* and finally the list of further subsections beginning with *Burgruine Werdenfels*, and *Historische Ludwigstraße*. A sequential menu is particularly effective in terms of cohesion on a web site, as current location and path travelled are indicated along with possible options for accessing further content.

Some web sites combine the global and local navigation functions in one where the

navigation tabs are at the same time part of a drop-down menu, as is the case in Figure 3.24 from the Leipzig site:



Figure 3.24: Combining global and local navigation, Leipzig site
(www.leipzig.de)

The main sections of this site are *BürgerPortal*, *BusinessLounge* and *TouristService*, as seen in the red, green and blue tabs respectively. The *TouristService* section is moused-over, showing its sub-sections *Leipzig*, *Szene*, *Unterkunft*, *Angebote*, *Kontakte* and *Presseportal*.

3.6.1.3 Embedded navigation systems - contextual

Contextual navigation refers to navigation that does not easily fit into the global or local navigation categories and generally does not belong in the overall structure and content of the web site. Links in contextual navigation are specific to a particular page, document or object. Contextual navigation allows the web designer ‘to create a web of connective tissue that benefits users and the organization’ (Rosenfeld and Morville 2002 p117). Links in contextual navigation could include a *See also* section or *Related topics*.

Figure 3.25 shows an example of contextual embedded navigation as a list of links with the common heading *What's on in Cologne*:



Figure 3.25: Contextual navigation, Cologne site
(<http://www.koeln.de/en/index.html>)

This example has links to additional information on related topics and events such as the *MusikTriennale* and the *Summerjam*.

Contextual links include links that are used to divide up a lengthy text, enabling the user to click a link and go directly to a particular part of the text. One such example can be seen on the page about Schloss Nymphenburg on the Munich site. The full text is divided into the five sections labelled *Nymphenburger Schlosspark*, *Amalienburg*, *Badenburg*, *Pagodenburg* and *Magdalenenklause*. The link labels

correspond precisely to these section headings enabling the user to go directly to each section:



Figure 3.26: Contextual navigation, Munich site

(http://www.muenchen.de/Tourismus/Sehenswuerdigkeiten/Burgen_und_Schloesser/132724/Schloss_Nymphenburg.html#Schlosspark)

Contextual navigation can include what is sometimes referred to as *inline* or *embedded* links. An embedded link is one that is placed within a piece of prose, either in a sentence or a paragraph, and is part of the text flow:

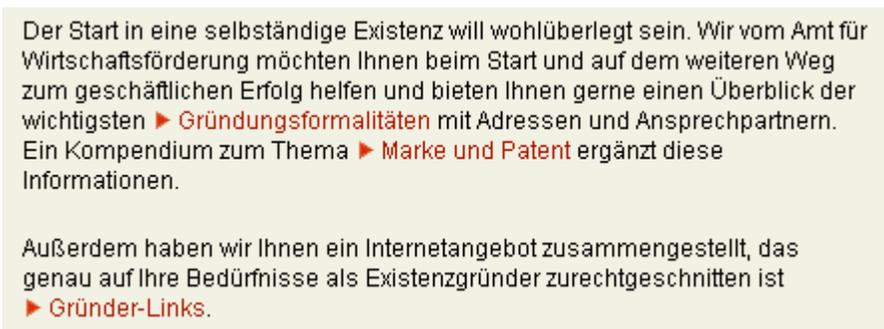


Figure 3.27: Contextual links, Regensburg site

(<http://www.regensburg.de/wirtschaft/existenzgruender/index.shtml>)

An embedded link takes the user to information that is of direct relevance to the piece of text in which it is embedded. This content can be on the same or on another web site. This destination content is, however, not necessarily directly related to the subject matter of the web site as a whole (Rosenfeld and Morville 2002 p117).

3.6.2 Supplemental navigation systems

Supplemental navigation systems are external to the general navigation hierarchy of a web site. They offer an additional means of finding and accessing content (Rosenfeld and Morville 2002 p121). Three common types of supplemental systems are the sitemap, the site index and the site-specific search feature.

3.6.2.1 Supplemental navigation systems - sitemap

A *sitemap* is a 'page which contains an organized listing of links to all pages within the website.' (SLS Consulting 2007). A typical *sitemap* will indicate at least the top levels of information hierarchy on a web site, providing a broad view of the content in the web site and facilitating random access to segmented portions of that content (Rosenfeld and Morville 2002 p122). A well-structured sitemap has an additional advantage in terms of online search as it facilitates crawling and indexing of the site (SLS Consulting 2007). The individual elements of a site map can be textual or graphic and if they are active links they provide the user with direct access to the page in question. The sitemap is a significant cohesive device on a web site as it

helps to reinforce the structure of the site for the user and facilitate fast and direct access to content, as seen in Figure 3.24 below:



Figure 3.28: Sitemap, Frankfurt site

([http://www.frankfurt.de/sixcms/detail.php?id=stadtfrankfurt_eval01.c.123086.de&_myvars\[_id_liste_nartikel\]=stadtfrankfurt_eval01.c.1707045.de](http://www.frankfurt.de/sixcms/detail.php?id=stadtfrankfurt_eval01.c.123086.de&_myvars[_id_liste_nartikel]=stadtfrankfurt_eval01.c.1707045.de))

It is of particular help for users who know what they are looking for (Rosenfeld and Morville 2002 p122) but do not have the precise name. The label text is often drawn from the page titles that are defined for the individual pages of the web site providing addition cohesion, for example between the labels of a sitemap and the page titles.

3.6.2.2 Supplemental navigation systems – site index

A *site index* is comparable to the index at the back of print material. It presents 'keywords or phrases alphabetically, without representing the hierarchy' (Rosenfeld and Morville 2002 p123). Site indexes are useful for people who have the exact name of what they are looking for. In this case, the user is less concerned with the

location of the required content in the overall hierarchy and more concerned with accessing the content:

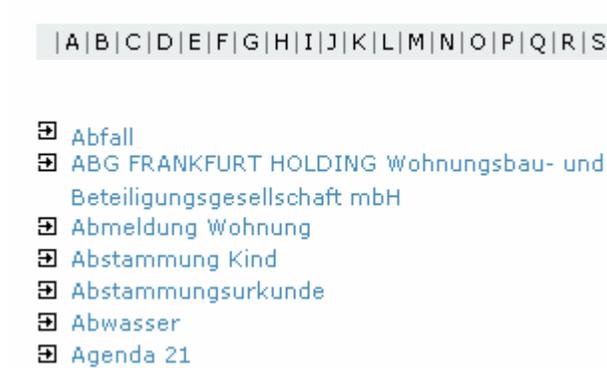


Figure 3.29: Site index, Frankfurt site

([http://www.frankfurt.de/sixcms/detail.php?id=2727&_myvars\[_id_listenartikel\]=219840&_searchvars\[title\]=A*#219840](http://www.frankfurt.de/sixcms/detail.php?id=2727&_myvars[_id_listenartikel]=219840&_searchvars[title]=A*#219840))

Some site indexes are not simply an alphabetical list of all items of content available on the web site but are sub-divided. The following example from the Frankfurt site indicates six sub-sections of the section *Stadtverordnetenversammlung* which are not arranged alphabetically:



Figure 3.30: Site index with sub-sections, Frankfurt site

([http://www.frankfurt.de/sixcms/detail.php?id=2727&_ffmparf_id_inhalt\]=12152&_ffmparf_verw\]=&_searchval\[re13093\]=&_searchval\[title\]=S*#12152](http://www.frankfurt.de/sixcms/detail.php?id=2727&_ffmparf_id_inhalt]=12152&_ffmparf_verw]=&_searchval[re13093]=&_searchval[title]=S*#12152))

Rosenfeld and Morville suggest that the sitemap ‘reinforces the hierarchy and encourages exploration, while the site index bypasses the hierarchy and facilitates known-item finding’ (2002 p123).

3.6.2.3 Supplemental navigation systems – site-specific search feature

A *site-specific search feature* is considered to be central to supplemental navigation and a favoured tool as the user him/herself decides what to search for and how to formulate the search terms (ibid. p127). Nielsen and Loranger consider it to be one of the most important design elements on a web site as users have come to rely heavily on internal search features (2006 p138). As can be seen in Figure 3.31, the Frankfurt site offers all three supplemental navigation features mentioned here: site map, site index and site-specific search feature:



Figure 3.31: Site-specific search feature, Frankfurt site
(http://www.frankfurt.de/sixcms/detail.php?id=stadtfrankfurt_eval01.c.317693.de)

The results of a site-specific search can be presented in different ways depending on

how the results are defined by the individual webmaster, as illustrated in Figures 3.32 and 3.33:

Suchergebnisse

↓ **Erweiterte Suche**

Es gibt 15 Treffer bei Ihrer Suche nach **restaurants** (Anzeige 1 - 15)

1. **Frankfurt trauert nach Niederlage im Halbfinale**
...nige Zuschauer verließen die nicht ganz ausverkaufte Arena schon vor dem Abpfiff, andere blieben fassungslos sitzen oder fielen sich weinend in die Arme. In Pizzerien, Eisdielen und italienischen **Restaurants** brach schon nach dem 1:0 der Jubel los. Bis in die Morgenstunden fuhren italienische Fans hupend und Fahnen schwenkend durch die Straßen. In mehreren Städten, darunter auch in Frankfurt, bildeten s... ..

2. **Lifestyle**
... Freizeit und Erholung lassen sich leicht miteinander verbinden. Die Entfernung zur Zeil, einer der bekanntesten und umsatzstärksten Einkaufsmeilen in Deutschland, zu den vielen Theatern, Kinos oder **Restaurants**, zu den schönen Parks oder zur reizvollen Natur des nahen Taunus und Odenwald ist in der Regel nicht weit. Auch Sportbegeisterte kommen in Frankfurt nicht zu kurz. Ob Fußball

Figure 3.32: Site-specific search result, Frankfurt site

(http://www.frankfurt.de/sixcms/detail.php?id=stadtfrankfurt_eval01.c.123086.de&_ffmpar_az%5B_gsid_inhalt%5D=stadtfrankfurt_eval01.c.1806768.de&_sprache=de&_searchval%5B_vtSearchmode%5D=fulltext_all&_searchval%5B_vtSuchwoerter%5D=restaurants)

8 Seite(n) mit dem Wort **'Restaurants'** gefunden.

[Alte Schmiede Altefähr](#)
[barrierefrei](#)
[gewusstwo](#)
[gewusstwo](#)
[Hotel Schweriner Hof](#)
[LaNa05](#)
[Stralsund putzt](#)
[Umfassendes, landesweites Portal zu den Themen Gesundheit, Ernährung, Wellness, Fitness, Tourismus](#)

Figure 3.33: Site-specific search result, Stralsund site

(http://www.stralsund.de/hst01/content.nsf/docname/Webseite_BE26DAAC95964232C1256E4B004A6840?OpenDocument)

In some cases, the results resemble a typical Google or Yahoo search result in that they carry the page title for the page to which they link and possibly the description content metatag for that page. Other site-specific search results consist of a link to the destination content.

3.6.3 Labels for navigation systems

Cohesion in navigation systems is created by ensuring that the link labels are capable of creating a cohesive link through repetition with, for example, a keyword in the heading of the destination page or document. Rosenfeld and Morville (2002 p92) refer to link text as a *label*, a label being a form of representation, indicating the presence of additional information.

The goal of a label is to communicate information efficiently; that is without taking up too much of a page's vertical space or a user's cognitive space (Rosenfeld and Morville 2002. p76).

They consider labels to be

as integral to an effective web presence as any other aspect of (a) web site, be it brand, visual design, functionality, content, or navigability. (ibid. p80).

Labelling is all the more important in the 'disintermediated medium' that is a web site, where the web site acts as message translator 'from owners and authors to users and back again' and where there are 'no visual or other cues, communication is harder' (ibid. p 77). Labels should be seen as part of a system rather than individual pieces of text that are or should be directly related to the content accessed when the link is clicked (ibid. p76). Expanding on this, Buchholz suggests that

though labels appear to exist independently (they) are inextricably bound to content, and together with content make up the information system. (2007 p2).

Nielsen and Loranger (2006 p168) consider good labelling to be important in the context of architectural SEO, and therefore, as part of the overall cohesive network, as link labels indicate to the web spider ‘what’s coming next’. In terms of SEO, keyword links on a web page are more valuable than simple links as they provide valuable ‘incoming’ links for the destination page which are highly rated by Google.

3.6.3.1 Embedded navigation – the labels

As a cohesive device, a label offers a precise, concise and unambiguous indication of the destination content in the language of that content, enabling the user to pre-suppose the destination content. The presence of foreign-language content on a web site is indicated on each page of a site, so that the user has at all times access to this content in the required language no matter what the page of entry to the site.

Where the destination content has been adapted in some way, that is to say updated, edited or abbreviated in some way to suit the TL audience, the SL label may no longer be an adequate indicator of the current content and will need to be adapted and translated accordingly. Labels are unique unless more than one label on a page or on a number of pages link to identical content.. Where a newly-coined term has to be used for a label, the meaning and therefore the nature of the destination content should nonetheless be clear.

Labels that are grouped together have a similar register, level of specificity and comprehensiveness. They also add to overall cohesion where they display parallel phrasing, which refers to starting each line and therefore each link label with the same type of word, i.e. a verb or a noun.

They reflect the current position in the overall hierarchy of the site, for example labels used to describe the main section of a web site will normally be more general than labels used for the sub-sections of that section, as the information provided in the sub-sections is generally more specific.

As mentioned in the previous chapter, keyword labels are considered to be high-value labels both from a user and a search engine point of view. It is important not to simply 'translate' the labels without due consideration given to the content of the destination page and the use of terminology that has already been defined for that page and positioned strategically in metatags and body text as potential cohesive devices.

Where possible, generic labels such as 'click here' or 'continue' are avoided as this would amount to a lost opportunity to create a valuable keyword link. As with all types of text on a web site, the most important information is placed as far left in a label as possible in order to increase the likelihood of this part of the text actually being read and acted upon by the user. Labels are sufficiently salient in comparison with non-hypertextual elements on a web page, which will facilitate cohesion when the user is scanning a page.

In the case of contextual links, the labels are meaningful in their immediate context although their meaning may not be related to the broad subject area of the web site in question.

3.6.3.2 Supplemental navigation – the labels

Despite the fact that the sitemap and the site index are important features on web sites today, the majority of the web sites in the research corpus do not use either. The decision was made, therefore, not to include these features in an analysis but to concentrate on the internal, site-specific search feature with which each of the web sites is equipped.

The site-specific search feature is placed in a prominent position on each web page and it is clear that the feature relates to a search on the site. Increasingly, however, web sites offer both possibilities side-by-side. There are a number of different formats for the list of search results with a site-specific search feature. However, the most important aspect of this is that the individual results are capable of setting up a cohesive link, through repetition, back to the search query. Ideally, the results returned for a query are only in the language of the user. With search results in a language other than the language of the user and of the search query, a cohesive link cannot be created between the TL term in the search query and the SL terminology in the search result. Some web sites display search results similar to those of the main SEs and using the text that has been defined and optimized for the relevant HTML tags and metatags

3.7 Navigation - applying the model

The analysis of cohesion in Navigation takes as its starting point the approach adopted by Baldry and Thibault (2006 pp153-4) to the analysis of web site content. They describe the objects in terms of different layers, where layer 1 is the visual

composition of the homepage, layer 2 is the text that appears when the object on the homepage is moused over and layer 3 is the page to which the object links. The text in layer 2 is anaphoric to the homepage and cataphoric to the next page. With relation to text, such anaphora and cataphora on a web site can only be lexical and not grammatical.

Similarly, the model is applied to Navigation involving a walk-through of the site. The first involves examining cohesion in terms of access to TL content via the global embedded navigation system. This involves not only ascertaining whether TL content is indicated across the web site but also how user movement within the TL content is facilitated by cohesion. In addition to considerations of access to the English-language content on the web sites, access to content in other languages was also examined in brief.

In order to explore cohesion involving link labels, a walk-through method is used to create a reading pathway, with the emphasis on content that is likely to be relevant to the tourist in the cities. The use of repetition as a cohesive device is investigated along such pathways.

In the case of the site-specific search feature, a search is carried out on each of the web sites using the TL search query 'museum'. The resulting search results were examined in terms of a) a cohesive link through repetition, linking search query, search result and destination content and b) the use of the SL language in search results.

3.8 Page Content – an overview

The third key area identified in Chapter II as presenting a challenge in terms of textual cohesion on the translated web site is *Page Content*. Page content refers to all on-page textual content of a web site, apart from the elements of navigation treated in the previous section. It includes the continuous text on a page in addition to features such as teaser text, widgets and information boxes. This study is concerned with elements of Page Content that are not classified as continuous text.

The homepage, main and sub-section pages of a web site are frequently navigational in character as they serve as a stepping stone to the actual content that the user wishes to access. While some of the items of text such as the teaser text consist in part of hyperlinks they are not considered to be solely navigational in nature for the purposes of this study and are therefore treated in this section as part of the general page content of a web site. A further example here is the widget which requires information from the user before it can provide information to the user. In this sense, while there is a hypertextual element to the widget, it is not considered to be a navigation tool *per se*. The three features of the navigation page to be addressed in this model are the *teaser text*, the *widget* and the *information box*.

3.8.1 Teaser text

A teaser text on a web site is a short, hyperlinked piece of text used to indicate to the web user the presence of additional content that for reasons of space or subject matter cannot be carried on the same page. The main objectives of the teaser, also referred to as a ‘blurb’, are, therefore, to attract the attention of the user, to awaken his/her curiosity and ultimately to persuade the user to click and access the destination content. While visual cues attract attention, it is generally the linguistic indicators that will persuade the user to read on and/or to click and access the destination content. Figure 3.34 shows a group of six teaser texts from the German homepage of the Munich site:



Figure 3.34: Teaser texts, Munich site
(<http://www.muenchen.de/>)

The teaser text consists of a number of individual items, as seen in the first of the six teasers in Figure 3.30: a heading (Da sein für München), a body text (Über den Aktionstag der städtischen Dienstleister am 13. 10...) and an icon link (☰). Each of the teasers on this page is accompanied by a supporting image. Some or all parts of the teaser are active links. In the above examples, with the exception of the body text all parts of the teaser are active links. Teaser texts vary considerably in length and can range from about 5 to 50 words. The teaser text is, potentially, a strong and effective cohesive device on the translated web site. It can be compared with the newspaper headline which Morley describes as a 'powerful producer of cohesion in newspaper discourse' (2006 p268). Cohesion is based in part on predictability where the headline, and by definition the teaser text, 'creates expectations about the contents of the article, its style, lexis and point of view, among other things' (ibid.). Consequently, a teaser text provides a precise and accurate description of the content to which it refers and contains keywords used in this text, thereby facilitating cohesion. In order to function fully as a cohesive device, a link must be created through repetition between the teaser text heading and body text, between body text and destination content and between heading and destination content.

The destination content of the teaser can be a single document, for example a PDF file. It can also be a content page, which is the case with the teaser text with the

heading ‘Da sein für München’ (above), which links to the page illustrated in Figure 3.35 below:



Figure 3.35: Destination page for teaser text in Figure 3.30, Munich site
 (<http://www.muenchen.de/Rathaus/dir/stadtspitze/obkolumn/39249/index.html>)

A teaser can also link to a navigation page containing additional teaser texts or link groups. In this case, the information supplied in the body text of the teaser can refer either to the general content of the destination page or it can refer to the subject matter of one of the destination teaser texts. Where the teaser links to a multi-topic page but is referring only to one of the topics on that page, the content to which the teaser refers is in a prominent position on the destination page, where possible and not, for example, below the fold as users may not scroll if they do not immediately find the information indicated in the teaser. Alternatively, the teaser text for such a multi-topic page can have a more general reference focus that manages to somehow encompass all topics mentioned on the destination page. The teaser labelled ‘Ausstellungen’ on the German homepage on the Munich site as seen in Figure 3.34

is one such example. Its destination page contains the following teaser texts, all of which can be categorized using the generic term ‘Ausstellungen’:



Figure 3.36: Destination page of teaser text ‘Ausstellungen’ in Figure 3.34, Munich site (http://www.muenchen.de/Stadtleben/Kultur_Unterhaltung/Museen_Ausstellungen/ausstellungen/106398/index.html)

Cohesion between these two pages is created through lexical (Ausstellungen, München), visual (Beckmann’s picture ‘Quappi in Blau und Grau’) and spatial (Beckmann picture in the same position among six teaser texts) links.

The teaser heading is of particular significance in the light of Eyetrack III, as discussed in Chapter II, which found that ‘users tend to scan a page, reading dominant headlines first’ (Outing and Rual 2004). The study also found that the user spends all of one second looking at the heading in a teaser, something which has considerable implications for the text in terms of providing as accurate an indication of the destination content as possible in as concise a manner as possible (ibid.). The

teaser header contains one or more keywords which, however, do not apply to the page on which the teaser is located but refer to the content of the destination page or the destination document. As the teaser heading is generally a hyperlink, the inclusion of a keyword in the heading has the added advantage of it being a valuable keyword link.

The body text of the teaser can be a summary of the destination content which has been specifically written for inclusion in the teaser text or it can consist of the first few lines of the destination text. It could also be a general reference to the subject area or topic of the destination page. It can happen that the destination content has been revised for a TL audience. Therefore, the teaser will need to be brought in line with this content, otherwise cohesion cannot be created between the two.

The Eyetracker Study III (Outing and Rual 2004) found that not only should blurbs or teaser texts be kept short but that the first few words of the blurb need to attract the attention of the user: ‘...most people just look at the first couple of words -- and only read on if they are engaged by those words’. It also discovered that very often only the left-third of a blurb or teaser text is actually read, provided the user reads beyond the headline at all. The following two examples show what it is the user sees in the case that

s/he a) only looks at the left third of the blurb and b) only reads the first line of the blurb:

Unterkunft in Berlin



In unserem Angebot findet sich alles vom eher preiswerten Unterbringung über das durchgestylte Design Hotel bis hin zur Edel Suite in einem der großen Top Hotels.

Unterkunft in Berlin



In unserem Angebot findet sich alles vom eher preiswerten Unterbringung über das durchgestylte Design Hotel bis hin zur Edel Suite in einem der großen Top Hotels.

Figure 3.37: Teaser texts displaying possible text read, Berlin site
(<http://www.berlin.de/tourismus/>)

While the translated teaser text may not naturally carry the most important words at the beginning of the text, the main keywords should nonetheless be placed there, even if it goes against syntactical conventions. If the teaser text is to function as a cohesive device on the translated web site, the destination content must be in the TL. Cohesion is facilitated by the salience of the visual cues that are part of the teaser text, for example the teaser heading in bold. However, as seen from the findings of Eyetrack III, a heading that is slightly removed from the main body of the teaser will not encourage the user to read on after the headline.

The Eyetrack study also found that users were more likely to read both the heading and the body text of the teaser if both texts had the same font size and there was no gap or line between them. Where the heading was larger than the body text of the teaser or separated by a shift space, no matter how slight, the user was less likely to carry on and read the body text. Where the teaser text is not adequately prepared for its function as a cohesive device, it may fail in its objective to draw to the attention of the user the presence of additional, relevant information on the site.

3.8.2 Widgets

The term ‘widget’ was created from the words ‘window gadget’. It refers to the elements of a graphical user interface, which is an ‘end-user interface to a program that presents information using windows, menus, buttons, text fields, etc., and with which the user interacts using a mouse or other pointing devices in addition to the keyboard’ (UCC CS 2007). The widget is a quick and easy means of ordering and paying for goods and services online. Hotel and flight booking forms are examples of some of the most popular widgets in use today. It is important that the user, when filling out these forms or clicking on the options provided, knows exactly what is expected of him/her. Widgets consist of data-entry fields, where the user inputs the required information him/herself, and/or of selection fields, where the user chooses

one of a number of options by clicking on it. The following widget is the German hotel reservation form available on the Berlin web site:

The screenshot shows a hotel reservation form with the following elements:

- Anreise:** 12, 10, 2007
- Abreise:** 10, 12, 2007
- Einzelzimmer:** 1
- Doppelzimmer:** (empty)
- Hotels ab Sterne:** 0
- Preislimit:** (empty) €
- Personen:** 1
- Search Button:** HOTEL SUCHEN (with a bed icon and arrow)
- Footer:** Datenschutz, Sicherheit, Impressum, AGB, Kontakt, Über HRS
- Search Bar:** Berlin search

Figure 3.38: Hotel reservation widget, Berlin site
(<http://www.berlin.de/tourismus-unterkunft/hotels/>)

This widget contains selection-based fields, such as the drop-down menu used to indicate arrival and departure dates (Anreise/Abreise) and input-based fields such as the boxes in which the number of rooms required are entered (Einzel/Doppelzimmer). A widget is capable of functioning cohesively where the labels provided for all data input and selection fields are accurate and unambiguous and in the correct language. Equally, the data option labels should be unambiguous, accurate and in the correct language, leaving the user in no doubt as to the information required from him/her when selecting options in the widget. Consistency is fundamental to cohesion in a widget. Where possible, labels for the data input and selection fields should be worded consistently so that the same data item has the same label if it appears on different pages (USDHHS 2006c). The analysis of widgets is user-oriented and involves the activity of booking accommodation using the online hotel booking form

on each web site. The terminology used in the widgets both for the field labels and the data itself is examined for clarity, unambiguity and language. The analysis does not include the actual booking of accommodation.

3.8.3 Information boxes

An information box consists of a small graphic image containing information in brief format for the user on a specific subject. Figures 3.35 and 3.36 show the German weather box from the Munich site and the Stock Exchange and Frankfurt Airport boxes from the Frankfurt site:



Figure 3.39: Information box, Munich site (www.muenchen.de)



Figure 3.40: Information boxes, Frankfurt site (www.frankfurt.de)

An information box is generally positioned on the margins of the page. However, in the case of Frankfurt these boxes are beneath the horizontal global navigation bar. Some information boxes have links to further information on the given topic while others do not. Information boxes can link to internal content or to external web sites such as the examples given in Figures 3.39 and 3.40 above. The Munich box contains three items (graphic image of the sun and the clouds, the link 'Weitere Werte' and the link 'wetter.com') that link to the same external meteorological web site at www.wetter.com: the. The links in the 'FrankfurtAirport' information box link to the real-time arrivals and departures timetable for Frankfurt airport on the official airport site.

3.9 Page content - applying the model

The application of the model in the case of the items of page content outlined as being an integral part of this section was once again user-oriented. The headings, body text and teaser text destination text of the teaser texts were examined for the presence of keywords in repetition and therefore cohesion. The English homepage and tourist-related pages were examined.

The part of a teaser text that would be visible if the user read only the left third of the text or the first line was examined for the presence of keywords that could create a cohesive link through repetition with the teaser heading.

The analysis of the widgets was user-oriented and involved an attempt to book a hotel room. Field labels, for example the label used for the list of different room types in a hotel, were examined in terms of a superordinate-hyponymic relation

with the individual items in the list. Widget text was also examined for the presence of non TL-text, which would prevent the creation of a cohesive link.

In the case of the information boxes, where they offered a link to further information the link was clicked and the destination content compared in language and content with the information provided in the original link.

3.10 Conclusion

This chapter presented a model for the analysis of cohesion on web sites. It addressed three key areas: *Online Search*, *Navigation* and *Page Content*. An overview of the function and significance of each of these sections was provided and the manner in which cohesion functions in each was outlined. The application of the model to the corpus web sites was described.

This model is applied to the corpus web sites and the results of the analysis of Online Search, Navigation and Page Content are presented in Chapters IV, V and VI respectively. The findings presented are not exhaustive but an indicative sample of what can be considered to be the most salient issues from the point of view of cohesion and the translated web site. As the issues that arise in the context of ‘information boxes’ are similar to those treated as part of ‘Navigation’ in Chapter V, they will not be addressed as part of Page Content in Chapter VI. The issues that arise are outlined and the implications for the translator are discussed.

Recommendations for dealing with the issues that arise in each of the analysis chapters are proposed in the Conclusion to this dissertation and are accompanied by a checklist as an aid to creating a cohesive web site in translation.

Chapter IV

Analysis of Cohesion in Online Search

Chapter IV

Analysis of Cohesion in Online Search

4.0 Introduction

Chapter IV presents the findings of the analysis of 'online search'. Three areas are outlined and discussed: the definition and use of keywords, the page title and the description content metatag. The implications for the translator are discussed.

4.1 Online Search

Online Search is an integral part of the Web experience for the great majority of Web users. It is a keyword-based search that seeks to match user requirements, as expressed through the search query, with web sites that have corresponding content, also expressed through the use of identical or similar terms. The cohesive link comes about through simple repetition between the search query term, the search result which contains that term and the web site to which the search query links. The positioning of these keywords on the web page and, in particular their inclusion in the HTML title tag and description content metatag is vital for the creation of a cohesive link. Cohesion is further enhanced by the inclusion in the search result of a precise and concise description of the page content. The most important aspect of the keywords, however, is that they reflect user language for the given subject matter.

The following findings are of particular significance in terms of cohesion and the translated web site:

- Online searches in English on www.google.ie using only the main keyword for each of the web sites are very successful, returning, without exception, links among the top three results of the SERP to each of the sites. Search queries using a keyword phrase, which contains the most important keyword for each site (the city name) and one other term (e.g. 'hotels') are not as successful as the queries using only the main keyword.
- Page titles defined for the English-content pages are at times inadequate. They do not contain the main keywords for the page for which they are defined and they fail to provide a true reflection of the page content.
- The description content metatags have not been defined for many English-language pages on many of the corpus web sites and where they have been defined they are frequently inadequate. In some cases where a description content metatag has been defined Google does not include that text in a search result but creates its own 'description'.

4.1.1 Analysis of keywords

In line with Nielsen's advice to 'speak the language of the user' (2006b), the analysis of the keyword as a cohesive device in online search adopts a user-oriented approach. The starting point, therefore, is not the web site and keywords that have been defined for its individual pages, but the terms that are used by the web user in a search query in the context of these cities. As mentioned in Chapter III, the most

popular keywords common to all 10 cities in the corpus were established using the software program, GoodKeywords. Searches were carried out on www.google.ie using firstly the main keyword for each site and secondly this keyword in conjunction with four other keywords: 'hotels', 'restaurant', 'entertainment' and 'things to do'.

4.1.1.1 Search using main keyword

The main keyword for all pages of each of the sites in the corpus is by default the name of the city. A search using the city name consistently returns a link to the web site on page one of the SERP. In the case of the Munich site, which will be used here as an example of a main keyword search, a link to the English homepage was in top position on the SERP and a link to the main tourist page on the site was in position two.

One reason for this high ranking is the number of incoming links from other web pages both on the same site and on other web sites. The number of incoming links to any web page that has been indexed by Google can be established using the Google search feature. As seen in Figure 4.1 below, a search using the tem 'link' followed by a colon and the URL for the page under investigation returns information on the

the number of links to the page in question. In this case, it is seen that the Munich English homepage has 5,830 incoming links:



Figure 4.1: Number of incoming links to the English homepage on the Munich site
(<http://www.google.ie/>)

This high ranking is also due to the strength of the city name as a keyword which results from the high level of usage of the city name in searches both as single queries and in combination with other terms. Any search for information on a particular city almost always includes the name of the city. For example, if the user is searching for a list of restaurants in Munich the term ‘Munich’ has to be included in the search query as otherwise a list of restaurants from the user locale will be returned in the SERP. Among the few exceptions here are, for example a search for unique objects such as ‘Reichstag’ or ‘Oktoberfest’ which are so well known that neither a city nor a country is required in order to return links to relevant web sites. A further reason for the city name being as strong as it is for all pages of each of the sites in the corpus is that, as a rule, the main keyword occurs frequently and naturally in the body text and in other positions on a web page. This creates almost by default a strong cohesive network that connects the user query via the search engine to the web site. The significance of the main keyword on a page is illustrated by the cached view in Figure 4.2 of a page from the Munich site, where the term

'Munich' is present as link text (Munich Tourist Office) and in the teaser texts

(Munich Sights; Munich Nights):



Figure 4.2: Cached snapshot of Munich tourist page with main keyword highlighted
(http://66.102.9.104/search?q=cache:Nf-N209VetoJ:www.muenchen.de/Rathaus/tourist_office/57799/+munich&hl=en&ct=clnk&cd=1&gl=ie)

The term 'Munich' is also present in the page title tag and the description content and keywords metatag for this page, see Figure 4.3.:

```
<title>muenchen.de - Munich Tourist Office</title>  
<meta name="description" content="Munich Tourism">  
<meta name="keywords" content="munich, tourist, office, hotel, oktoberfest">
```

Figure 4.3: Title tag and keywords and description content metatags for Munich tourist page
(http://www.muenchen.de/Rathaus/tourist_office/57799/)

Finally, with the exception of the terms Munich, Cologne and Münster, the main keywords for the web sites do not need to be translated or re-defined. A search for 'Munster' as opposed to 'Münster' returned a result in eighth position in the SERP, despite having to vie with the many web sites listed for the Irish province of Munster. The high ranking of the English pages returned for the single-term search

query reflects the success of identical searches on www.google.de where a result for each SL site in the corpus was returned, without exception in top place in the SERP.

4.1.1.2 Search using a keyword phrase

While searches using the city name as single keyword returned, without exception, links to the web sites in the corpus among the first ten in a SERP, searches using a keyword phrase consisting of the city name in conjunction with one other keyword produced mixed results. The analysis using a keyword phrase as search query is carried out with terms that regularly co-occur in search queries in combination with the main keyword for each of the cities in the corpus: ‘hotels’, ‘restaurants’, ‘entertainment’ and ‘things to do’, as established with GoodKeywords. An example would be ‘Munich hotels’. Links to some of the corpus web sites were returned among the first ten search results on the SERP, others were considerably further down the list. Some web sites were not listed at all by Google.

Of the ten searches carried out using the keyword phrase ‘city name + hotels’, a link to the English content of the Regensburg and Munich web sites was returned on page one and two respectively, while links to the English content of the Münster and the Cologne sites were each on page 4 of the SERP. The links listed for the Garmisch-Partenkirchen and Berlin sites had German destination content. Searches for content about Frankfurt, Hamburg, Stralsund and Leipzig returned no links to the corpus web sites. In comparison, similar searches in German on www.google.de returned links to the Munich, Münster, Regensburg, Garmisch-Partenkirchen and Cologne sites among the first ten listings on the SERP, while Frankfurt, Hamburg and Stralsund were not listed at all.

The search query ‘city name + restaurants’ on www.google.ie did not generate a link to any of the corpus web sites while the same search in German on www.google.de generated results for München, Köln and Berlin, all of which are listed among the first ten results on the SERP.

Searches for ‘city name + entertainment’ returned links only to Munich, Berlin and Münster, while searches using the query ‘city name + things to do’ generated no results for the corpus web sites.

4.1.2 Issues with keywords in Online Search

Searches using the city name as a single term query are exceptionally successful for all web sites in the corpus. This high ranking of the main keyword for each of the web sites is, however, not capitalised upon where the city name is combined with a second, popular keyword such as ‘restaurant’ or ‘entertainment’. There are a number of reasons for such unsuccessful searches. Firstly, while certain popular keywords may be present in the body text of a page, they will not be highly rated by Google if they are not included in what is perceived to be the most important places on a web site from a search engine point of view which is the page title, the page heading and in hyperlinks.

In addition, the web sites differ considerably in how they label, for example, their ‘Restaurant’ sections. Garmisch uses the label ‘Food and Beverage’, Münster refers to ‘Gastronomy’, Cologne has ‘Eat and drink’ and Frankfurt has ‘Restaurants and Shopping’. Munich, which will be looked at in more detail below, uses the phrase ‘Eating out’. This variety of terminology is not only evident from one web site to another but also at times within the same site where the page title and link labels and

perhaps even page headings have differing labels to describe the content on that web page.

The phrase ‘eating out’ is an example of the significance of user language in online search. ‘Eating out’ is a popular phrase in English and appears to be used frequently on English SL web sites. However, it would appear that the generic term ‘restaurant’ or ‘restaurants’ is more commonly used in online search, as seen in the GoodKeywords list in Figure 3.15 for the term ‘cologne’. This reinforces Nielsen’s recommendation to use ‘old words and simple words’ (2006b) when writing for the Web, as users tend to use the first term that comes to mind in a given context. This is often a term that has been in use for a very long time as opposed to a newer or more trendy term. The Munich web site contains a comprehensive list of restaurants, beer gardens, bistros and cafés in English, in addition to a glossary of culinary terms in German with their English translation. Nonetheless, a search for ‘munich restaurants’, as mentioned, returns no listing for this content. This is despite the fact that a search for SL content on www.google.de using the keyword phrase ‘restaurants münchen’ or ‘münchen restaurants’ returns a link to the German content of the same site in each case among the first ten results on the SERP.

The screenshot below shows the main page of the restaurant section in the English content on the Munich site. The keyword ‘restaurant’ appears in the body text of two teaser texts (Hofer – der Stadtwirt; Nektar) neither of which is a hyperlink.

However, the keyword is not present in the page title, which reads ‘muenchen.de – Eating out’. Nor is it present in the page heading which reads ‘Gastro “Eating Out” (Ads)’. The menu bar on the left in Figure 4.4 below contains link labels with the names of various establishments. The term ‘restaurant’ is not included:



Figure 4.4: Main restaurant page, Munich site, without keyword ‘restaurant’ strategically positioned
 (http://www.muenchen.de/Stadtleben/Eating_out/7860/index.html)

The cohesive network created in the parallel ST in German through the use of

repetition with the term ‘restaurant’ is illustrated in Figure 4.5, the cached snapshot of this page:

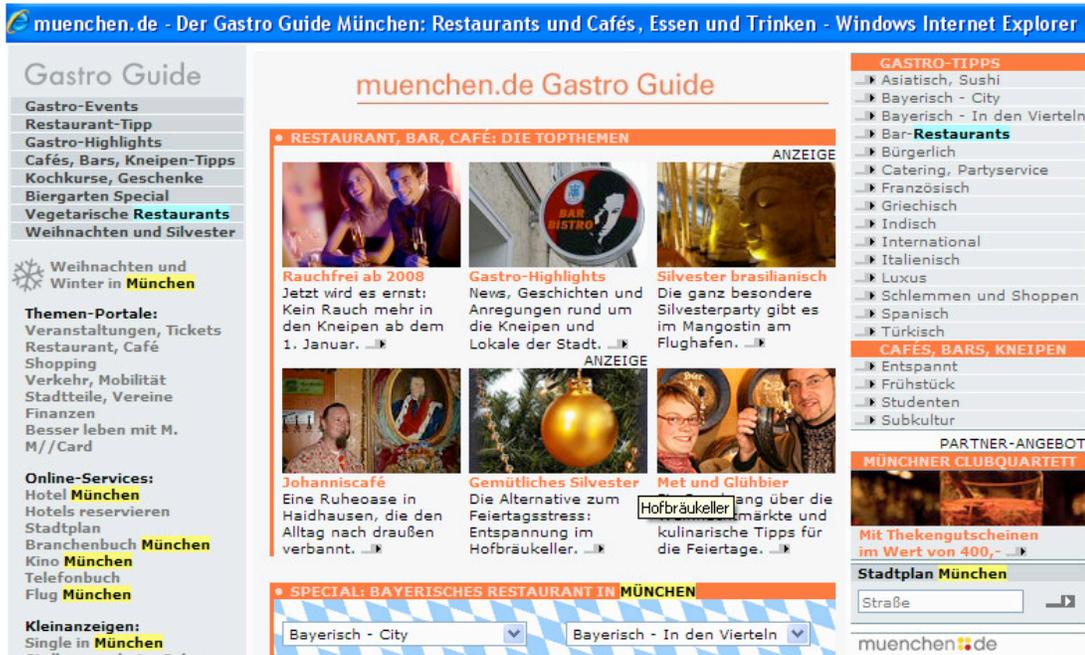


Figure 4.5: Cached snapshot of main restaurant page in German content, Munich site (<http://www.muenchen.de/verticals/GastroGuide/118073/index.html>)

In this example, the main keywords ‘München’ and ‘restaurants’ are included in the page title and as keyword links. The heading ‘muenchen.de Gastro Guide’ is a graphic element and is not recognised as text by the web spider.

A further issue with the Munich site in the search for English content is the use of the word ‘muenchen’ instead of ‘Munich’ in the page title. The search engine is seeking to match the text ‘munich restaurants’ with, for example, the page title in the English content but registers only ‘muenchen.de – Eating out’, where there is no lexical repetition. A cohesive link is not created between the search query and the web page.

This issue demonstrates the importance of including certain keywords and keyword phrases in the English page title, page heading and keyword links. It also points up the significance of including, for example the name of the city in the page title in the TL as this is the form of the word that a web user will most likely use when searching for information in the TL on that city. This is particularly the case given the spelling of a word like 'München' which is quite different to the English version in addition to containing an umlaut which the average English writer will not have on his/her keyboard.

A further issue that has a direct influence on cohesion in online search concerns the categorization of content items on the web sites and the subsequent labelling of these categories. The results of the searches using the keyword phrase 'city name + entertainment' illustrate this point. As mentioned above, searches with this keyword phrase generate links to the Munich, Berlin and Garmisch web sites only, despite the fact that with the exception of Regensburg, the sites have varying levels of content on different types of entertainment in their cities. It is assumed that the term 'entertainment' is understood to mean an event or performance designed to entertain. It is often but not always something that takes place in the evening and it does not necessarily require direct participation on the part of the person being entertained. Information provided about entertainment could conceivably include cinema, concert and theatre listings and nightclubs.

However, in many cases the word 'entertainment' is not included on the web site. The headings for the sections with content on entertainment differ from site to site and include 'What's on' (Cologne), 'Recreation/culture/sport' (Stralsund), 'Events'

(Frankfurt) and 'Culture and leisure' (Münster). In addition, the Leipzig site has a category labelled 'Culture and Eating out' which consists of a wide variety of sub-sections including: 'Music and theatre', 'Museums', 'Environment/Sports/Recreation', 'Shopping', 'Nightlife', 'Events' and 'Eating out (culinary survey)'. The Cologne site has placed items that could be included under 'entertainment' in three different sections. For example, 'What's on' includes 'Events and Highlights', 'Clubbing' and 'Cinema', the section 'Arts and Culture' includes 'Museums', 'Theatre' and 'Music' and the section 'Tourism' includes 'Events and Highlights' and 'Carnival'.

While the term 'entertainment' may not be an exact translation of the various terms found in the German texts such as 'Veranstaltungen', 'Events', 'Unterhaltung' and 'Kultur', to mention a few, it is a superordinate that is commonly used in online searches in English when searching for various types of entertainment.

The Berlin site has introduced the keyword 'entertainment' as a superordinate in the English content section. It is included as a menu tab, that is as a keyword link and in the page title of the main page in that section, two important positions for cohesion in online search.

The Munich tourism section menu bar, as seen in Figure 4.6 below, is an interesting example of how content can be categorized and labelled in such a way as to utilize the high ranking already enjoyed by the city name of the web site and facilitate

cohesion between user, search engine and the web page, to which the search result links:

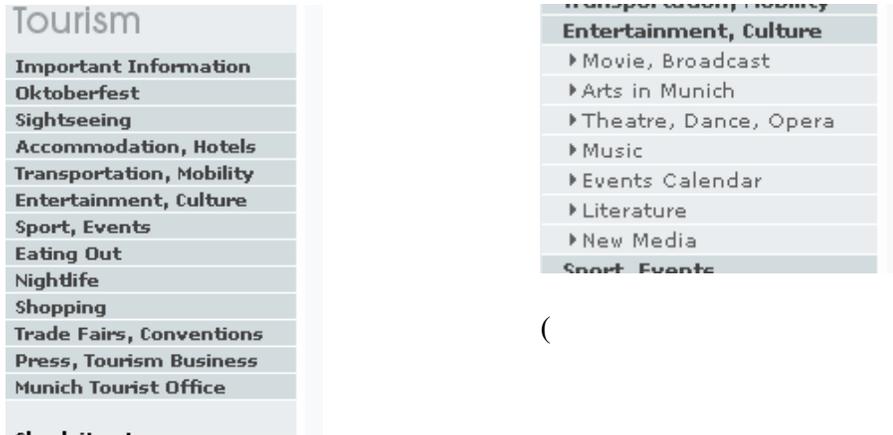


Figure 4.6: Menu bar 'Tourism' section with sub-section 'Entertainment, Culture', Munich site (<http://www.muenchen.de/Tourismus/6942/index.html>)

The English entertainment section on the Munich site includes various types of 'entertainment' and has been adapted to reflect tourist needs. 'Nightlife', which is found in the 'Jugend' section in the German-language content, is a section in its own right in the English content. The main page for the section 'Entertainment, Culture' has the page title 'muenchen.de – Entertainment, culture'. In addition, the word 'entertainment' is included as a keyword link in the menu bar on the left and is included in the h1 header all of which are important for the web spider and therefore for search engines.

As seen with the example of the search results, despite the presence of considerable content a cohesive link cannot be created if the cohesive device of repetition is not employed in the TT. The consistent use of 'old keywords' in the appropriate places as opposed to 'catchy' phrases such as 'What's on', as seen on the Cologne site would aid cohesion.

The implications are serious. The opportunity to deep-link and take the user directly to the content required and not simply to the homepage of a web site cannot be exploited. A more serious issue is where such content on a site is not found at all through online search due to the lack of cohesion.

4.1.3 Analysis of page title

Page titles appear to have been defined for most if not all pages in the English content on the corpus web sites. However, the format at times varies widely both within the same site and from one site to another. The issues addressed in this section include: page titles for TL content that are not in English, unique page titles and optimization of page titles for skimming and scanning.

4.1.3.1 Page titles in the SL

The page title defined for the English homepage on the Garmisch-Partenkirchen web site has not been translated:

```
<title>Garmisch-Partenkirchen - Urlaub in Garmisch-Partenkirchen - Ferien,  
Wandern, Skifahren, Wellness</title>
```

Figure 4.7: Page title, English homepage Garmisch-Partenkirchen site
(<http://www.garmisch-partenkirchen.de/en/db41673c-3301-78c8-8a27-52cc9b4d45f6.html>)

All other page titles on this site appear to be in English.

Some web sites in the corpus use a mixture of English and German in their page titles, the Munich and Regensburg web sites being two examples. The text 'muenchen.de' is included in all page titles on the Munich site, both English and

German. The following is the page title for the English homepage:

```
<title>muenchen.de - Munich, Germany</title>
```

Figure 4.8: Page title, English homepage Munich site
(<http://www.muenchen.de/home/60093/Homepage.html>)

Regensburg includes not only German in some page titles for the English content but also French and Italian. The title:

```
<title>tourismus.regensburg - Hotels - Accomodation - Chambres –  
Camere</>
```

Figure 4.9: Accommodation section, English content Regensburg site
(<http://www.regensburg.de/tourismus/international/common/accomodation/index.shtml>)

has been defined for the ‘Hotels & Accomodation’ (sic) page.

4.1.3.2 Unique page titles

The majority of web sites have defined a unique page title for each of the pages on their web site similar to the following pages from the Cologne web site:

```
Cologne – What’s on?  
Cologne – Cinema  
Cologne – Art + Culture
```

However, in some cases an identical title is used for a number of pages within the same section of a site or for all pages on a web site. Stralsund uses the same page title for all pages (Hanseatic Town Stralsund) and, with the exception of one page, Münster also uses the same page title (Stadt Münster: Münster Marketing), the exception being the events page (Münster Marketing).

4.1.3.3 Page titles not optimized for skimming and scanning

The consistent placement of the most important keyword at the start of a page title, followed by the secondary keywords, facilitates skimming and scanning and, therefore, enhances cohesion, particularly with history or bookmark lists.

The Frankfurt site begins each page title with the name of the city followed by a colon and a description of the individual page content:



Figure 4.10: History list, Frankfurt site

The Berlin site is similar to the Frankfurt site in that all pages include the name of the city, 'Berlin.de'. However, the Berlin site at times places this keyword at the start of the page title and other times at the end, as can be seen in the history list in Figure 4.11 below:



Figure 4.11: History list, Berlin site

Some page titles are particularly long, such as the title which has been defined for

the main page of the business section on the Regensburg site, as seen in the reverse bar in Figure 4.12:



4.12: Page title, international business section Regensburg site
(<http://www.regensburg.de/wirtschaft/international/index.shtml>)

While a page title is generally displayed in its entirety in the reverse bar at the top of the computer screen, this is not the case with a search result. In this case, approximately 50 characters of the page title will be displayed in the first line of the search result, as seen in Figure 4.13:



Figure 4.13: Page title in search result, Regensburg site

In the example in Figure 4.13, the main keyword (Regensburg) is not visible in the first line of the search result as it is placed at the end of the page title.

The Garmisch site shows a high level of consistency with the word order in its page

titles and includes the name of the city, followed by the name of the main section and then the sub-section, where applicable:



Figure 4.14: History list, Garmisch-Partenkirchen

However, as the name ‘Garmisch-Partenkirchen’ is quite long, the full text of the page title is not visible.

4.1.4 Issues with page titles

The lack of adequately-defined page titles in translation can have far-reaching implications affecting not only cohesion in online search but also the cohesive value of history and bookmark lists and reverse bars.

The omission of the main keywords from the page title leading to a lack of cohesion in online search was addressed in section 4.1.2.

A TL page that has a page title that is not in the TL can have serious implications for both the search engine and the user, as the search engine will be unable to match the terms used in a search query with the page title and the page content it has indexed and will be unable to return a link to the relevant page. Even where the search engine does succeed in returning a link to a specific page, it must be assumed that the user will be unable to understand the first line of the search result if it is in a

language other than the expected TL. The user will remain unaware of what may be relevant and valuable content. Equally, if the page title is in the SL, there can be no cohesive link through repetition in bookmark or history lists or in the reverse bar if it cannot be read by the user.

The lack of a unique page title for each page of a web site has serious implications for cohesion on a web site. Google will list up to two pages from the same web site in a SERP. If these pages have identical page titles, the user will not immediately know that each result links to a different page, as is the case with the search results in Figure 4.15 for the Münster site:

[Stadt Münster: Münster Marketing](#)
Information: **history**. A Brief **History** of **Münster**. 793: The Friesian Liudger, dispatched by Charlemagne to undertake missionary work, arrives at the Saxon ...
www.muenster.de/stadt/tourismus/en/history.html - 14k - [Cached](#) - [Similar pages](#)

[Stadt Münster: Münster Marketing](#)
Many impulses and associations spread from here revealing new aspects of **Münster's history** and special features, initiating dialogue and making the areas of ...
www.muenster.de/stadt/tourismus/en/city-guides_sculpture.html - 17k - [Cached](#) - [Similar pages](#)

Figure 4.15: Search results for Münster site

The user may have to read the description texts, in order to establish the content of the page to which the results link. While the description text may give a precise indication of the page content, this is not always the case. It can also happen that the user will decide not to read beyond the first line of the search result. Either way, it is confusing and unnecessarily detains the user.

The sense of context and therefore cohesion created by the reverse bar on the screen is seriously compromised where some or all of the pages on a web site have an identical page title, as the result is a virtual ‘Groundhog Day’. In addition, non-unique page titles on a site render the history list redundant, forcing the user to access each link in the list in order to establish the content of each individual page. The pages listed in Figure 4.16, all of which have different content and were accessed on the Münster site on the same day, illustrate this point:



Figure 4.16: History list, Münster site

The lack of unique page titles for all pages on a translated web site may have an effect on how the pages are indexed by the search engine as the Web spider may be unable to differentiate between two pages with the same page title. Consequently, web pages that are relevant to a search query may not be identified as such and will not be returned in a SERP.

4.1.5 Analysis of description content metatags

The description content metatag defined for a web page has the potential to function cohesively by providing the user an accurate and concise description of the page content for which it has been defined. The analysis of the description content metatag involved carrying out a search on www.google.ie for each of the web sites

using the city name as search query. The description text returned in each of these search results is analysed in terms of content, keywords and origin.

Searches using the city name returned all ten web sites on page one of the SERP, four of which were in first place. Six of the ten pages that are returned in the search results have a description text defined for the page listed and four have none.

However, only one of the six description texts that have been defined is used by Google in the search result. This is the description for the Cologne page. The other five are not used.

4.1.5.1 Google uses description content metatag

The search result returned for the query 'Cologne' links to the English homepage of the corpus site and displays part of the description content metatag that is defined for the page:

[Cologne - City online](#)

Between Kölsch and carnival - and with joie de vivre and flair into the bargain. Front page which gives an overview and links to further sites with ...

[www.koeln.de/en/ - 23k](#) - [Cached](#) - [Similar pages](#)

Figure 4.17: Search result, Cologne site

The search result displays the first 121 characters of the description content metatag which in full reads as follows:

```
<meta name="description" content="Between Kölsch and carnival - and with  
joie de vivre and flair into the bargain. Front page which gives an overview  
and links to further sites with information about the tourist attractions,  
history, cultural life, events/venues, routes on foot, sport, hotels/restaurants  
and excursions." />
```

Figure 4.18: Description content metatag, English homepage Cologne site
(<http://www.koeln.de/en/index.html>)

The second half of the description (researcher's italics) is not displayed in the search result. This section contains a reference to accommodation, tourist attractions and events. The description content metatag does not include the main keyword for the page, 'Cologne'.

4.1.5.2 Google uses snippets in place of description content metatag

Google at times chooses to display a description text that has been created from snippets of text, regardless of whether a description content metatag is defined for a page or not. These snippets are drawn either from text on the page that contains the search query or from a web page that is linked to the page which is to be displayed in the SERP.

The search result returned for the query 'Munster' consists of one single snippet of text taken from the body text on the English homepage of the Münster site:

[Stadt Münster: Münster Marketing](#)
In **Münster** you find everything you can wish for during an exciting short stay: A city with plenty of history as well as shopping opportunities and green ...
www.muenster.de/stadt/tourismus/en/ - 15k - [Cached](#) - [Similar pages](#)

Figure 4.19: Search result, Münster site

(<http://www.google.ie/search?hl=en&q=m%C3%BCnster&btnG=Search&meta=>)

It highlights the keyword 'Münster' despite the fact that the search query was 'Munster' which is not present either in the page content or in the HTML tags. The description metatag defined for this page is:

```
<meta name="description" content="Tourist Information of Münster (Westphalia)">
```

Figure 4.20: Description content metatag, English tourism page, Münster site

(<http://www.muenster.de/stadt/tourismus/en/index.html>)

The search query ‘Munich’ returned a result in first position in the SERP linking to the main ‘Tourist Information’ page in English:

[muenchen.de](http://www.muenchen.de) - **Munich** Tourist Office
Enjoy the holidays in one of **Munich**'s Royal Castles weiter · **Munich** Hotel **Munich** Nights ...
In 2007 there are many good reasons to visit **Munich!** weiter ...
www.muenchen.de/Rathaus/tourist_office/57799/ - 63k - [Cached](#) - [Similar pages](#)

Figure 4.21: Search result, Munich site
(<http://www.google.ie/search?hl=en&q=munich&meta=>)

A description metatag is defined for the English page to which the search result links although none was defined for the corresponding SL page:

```
<meta name="description" content="Munich Tourism">.
```

Figure 4.22: Description content metatag, English tourism page, Munich site
(<http://www.muenchen.de/Tourismus/6942/index.html>)

Unlike the description created by Google for the Münster search result, the one for the Munich search result uses three snippets taken from three different teaser texts on the page to which it links. Figure 4.23 is a cached photo of the page to which the search result links showing the snippets that are used in the description each of

which contains the search query ‘Munich’ (highlighted):



Figure 4.23: Cached snapshot of English tourism page, Munich site
(http://216.239.59.104/search?q=cache:Nf-N209VetoJ:www.muenchen.de/Rathaus/tourist_office/57799/+munich&hl=en&ct=clnk&cd=1&gl=ie)

The two instances of the German term ‘weiter’ in the search result above appear to be a conversion to text of the arrow icon “” which is used to access further information on the topic.

The search query ‘Hamburg’ generates a result that links to the English homepage:



Figure 4.24: Search result for Hamburg site
(<http://216.239.59.104/search?hl=en&gl=ie&q=hamburg>)

Once again, Google has created the description from snippets taken from the page to which the search result links:

Welcome to Hamburg



The inhabitants of **Hamburg** are proud to have a lake right in the heart of their city and at only a short distance from the Opera, the Kunsthalle Art Museum, the City Hall and the Stock Exchange.

They are convinced that they live in one of the most beautiful cities in the world. Whether this is true, is for visitors themselves to discover.

Hamburg looks forward to your visit.

Hamburg and the world

Hamburg's geographical location is so favourable, that it has always been within easy reach for the rest of the world. With more

Figure 4.25: Cached snapshot of English homepage, Hamburg site
(<http://international.hamburg.de/?ok=18338>)

A description metatag has not been defined for the English homepage of the Hamburg site, but one does exist for the German homepage:

```
<meta name="Description" content="hamburg.de - hier finden Sie alles über Hamburg: News, Veranstaltungen, Musicals, Hotels und Sport mit zahlreichen Online-Buchungsmöglichkeiten" />
```

Figure 4.26: Description content metatag, German homepage, Hamburg site
(<http://www.hamburg.de/>)

4.1.5.3 Google uses Web directory extracts for description

In many cases, the search engine draws on other web sites, in particular directories in order to create a description text for a search result. Many of the examples are

drawn from the Open Directory Project, DMOZ. The content in this directory is used extensively by Google and is enhanced using Google technology (Google Directory 2007).

The search result returned for the query 'Frankfurt' links to the English homepage:



Figure 4.27: Search result, Frankfurt site
(<http://www.google.ie/search?hl=en&q=frankfurt&meta=>)

The description metatag defined for the English homepage is:

```
<meta name="DC.Description" content="Frankfurt: to the most international city in Germany, the largest financial centre on the continent, the historical city of coronations, the city of Goethe and the Frankfurt School... In brief, to the smallest..." />
```

Figure 4.28: Description content metatag, English homepage, Frankfurt site
(http://www.frankfurt.de/sixcms/detail.php?id=stadtfrankfurt_eval01.c.317693.en&template=hp_flash)

This consists of the first few lines from the letter of welcome from the Lady Mayor of Frankfurt on the English homepage, see Figure 4.29:

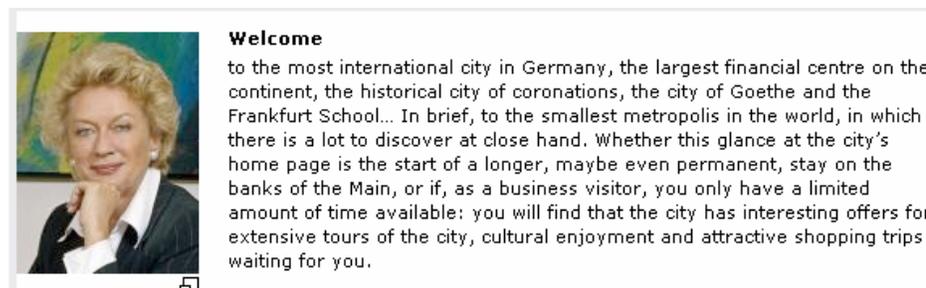


Figure 4.29: Body text, English homepage, Frankfurt site
(http://www.frankfurt.de/sixcms/detail.php?id=stadtfrankfurt_eval01.c.317693.en&template=hp_flash)

It is a departure from the text defined for the German homepage:

```
<meta name="DC.Description" content="Frankfurt: Die offizielle Homepage  
der Stadt Frankfurt am Main mit Informationen zum Leben in Frankfurt am  
Main, zu Tourismus, Kultur und Wirtschaft." />
```

Figure 4.30: Description content metatag, German homepage, Frankfurt site

(http://www.frankfurt.de/sixcms/detail.php?id=stadtfrankfurt_eval01.c.317693.de&template=hp_flash)

The search result generated for the query ‘Berlin’ is in first position on the SERP and links to the English homepage. The description text is similar to that returned in the Frankfurt result mentioned above:

[Information in English - Berlin.de](#)
Official guide to the history, people, and places of **Berlin**. Government, business, and tourist information. (German/English)
www.berlin.de/english/ - 32k - [Cached](#) - [Similar pages](#)

Figure 4.31: Search result, Berlin site

The description defined for the English homepage for Berlin is, at 240 characters, considerably longer than the text length preferred by Google these days:

```
<meta name="description" content="Welcome to Berlin's official internet  
site. Here you can find information about different aspects from Berlin:  
Hotels and Hotel Booking, Visitors Guide of Berlin, Entertainment in Berlin  
and Tickets, Political System and Local Authorities, Berlin Capital City,  
Business in Berlin" />
```

Figure 4.32: Description content metatag, English homepage, Berlin site

(<http://www.berlin.de/english/>)

It begins with the word ‘Welcome’ and contains the word ‘Berlin’ six times in a text of 40 words. This is a higher main keyword density than is the case with the ST which has the word ‘Berlin’ twice in a text of 28 words.

Searches with the queries ‘Regensburg’ and ‘Leipzig’ return results in second position on the SERP that link to the English homepage on each site:

[tourismus.regensburg - English Information](#)
Official information provided by the municipal government.
www.regensburg.de/tourismus/international/english/index.shtml - 26k -
[Cached](#) - [Similar pages](#)

Figure 4.33: Search result, Regensburg site

[Leipzig International](#)
Service for Citizens, Business Lounge and TouristService.
www.leipzig.de/int/en/ - 18k - [Cached](#) - [Similar pages](#)

Figure 4.34: Search result, Leipzig site

Neither of these description texts contains the main keyword. The Regensburg description text does not provide a description of the content of the page to which it links, which contains information for both tourists and business people visiting or living in the city. Neither of the sites has defined a description metatag for either their German or English homepage.

4.1.5.4 Description text in the source language

The result returned for the search query ‘Garmisch-Partenkirchen’ links to the German homepage. It displays the description text as defined in the metatag:

[Garmisch-Partenkirchen - Urlaub in Garmisch-Partenkirchen - Ferien ...](#)
Offizielle Website von **Garmisch-Partenkirchen** mit Infos zu Tourismus und Bürgerservice.
Quartiersuche mit Online-Buchung, Veranstaltungstipps, ...
www.garmisch-partenkirchen.de/ - 3k - [Cached](#) - [Similar pages](#)
[Hotels, Pensionen und ...](#) - www.garmisch-partenkirchen.de/de/008cbc86-9e6...
[Gastgeber & Hotels - Unterkünfte](#) - www.garmisch-partenkirchen.de/de/dbdfb...
[Marktplatz](#) - www.garmisch-partenkirchen.de/de/marktplatz.html?action=inde...
[Bürgerservice](#) - www.garmisch-partenkirchen.de/de/c19c16c1-90a3-41a4-e2fd...
[More results from www.garmisch-partenkirchen.de »](#)

Figure 4.35: Search result, Garmisch-Partenkirchen site

A description content metatag has been defined for the English homepage of the Garmisch site which is identical to that of the German homepage.

Similar to the situation with Garmisch-Partenkirchen, the description in the search result returned for the query 'Stralsund' is in German:



Webseite www.stralsund.de
Dies ist die offizielle Webseite der Hansestadt **Stralsund**.
www.stralsund.de/ - 4k - [Cached](#) - [Similar pages](#)

Figure 4.36: Search result, Stralsund site

The search result in Figure 4.36. links to the only page on the web site that has defined a description content metatag. That is the entry page for the German site which is an introductory page that contains no content and fades away after a few seconds.

4.1.6 Issues with description content metatags

Despite the fact that Google used the description text as defined for the Cologne page, there are a number of issues that need to be addressed in order to ensure that it can function cohesively. The Cologne homepage description consists of 244 characters, while Google appears to favour a maximum of two lines which equates to approximately 140 characters. As a result, more than half of the text will not be displayed in the search result. The text displayed in Figure 4.17 above (as seen

below in Figure 4.37) gives no indication of the actual content of the web page to which it links:

Between Kölsch and carnival... with joie de vivre and flair into the bargain.
Front page which gives an overview and links to further sites with ...'

Figure 4.37: Description text as seen in search result, Cologne site

The positioning of the above text at the beginning of the metatag is problematic as the information that is of relevance to the tourist is not displayed, as seen in Figure 4.38:

... information about the tourist attractions, history, cultural life,
events/venues, routes on foot, sport, hotels/restaurants and excursions.

Figure 4.38: Description text defined but not displayed in search result, Cologne site

The German word 'Kölsch' is possibly unknown to those who are not familiar with Cologne. The displayed text contains no keywords. At the very least, the main keyword, Cologne, should be included and placed at the start of the text, followed by secondary keywords, as appropriate. Saliency plays an important role here as keywords are highlighted in a search result, which facilitates scanning. The inclusion of keywords in the description content metatag is also important for the web spider and, therefore, for overall cohesion in online search.

The use of snippets to create a description for a search result is not always successful as the snippet rarely succeeds in providing an accurate description of the content of the page to which the search result links. As seen with the examples above, the snippets are relevant when read in context but make little sense when taken

randomly and placed in a search result. Though they may contain the main keyword for the page, they rarely provide a cohesive link to the web page due to the lack of description provided and the sense of optical fragmentation caused by the use of a number of snippets of text.

While it is not exactly clear what criteria Google uses when choosing to use the information provided in a description content metatag or not, the likelihood is greater that it will be used if it is approximately 140 characters in length, contains at least one or two of the main keywords for the page, does not repeat these keywords and does not include the word 'Welcome' in the text.

Likewise, the criteria by which Google decides to use a directory entry are not very clear. In the case of Frankfurt and Berlin, the directory extracts are more concise and accurate than the description text that has been defined for the pages. The main issue with the Leipzig and Regensburg directory descriptions is that the main keyword is not included which would be useful for scanning search results as the search queries are highlighted in the result. In addition, the description text in the Regensburg search result gives no indication of the actual content of the page to which it links.

A search for English-language content about Garmisch-Partenkirchen will not link to content in English as the language metatag contained in the HTML code is set to

default to German-language sites, as indicated by the text ‘de’:

```
<meta name="language" content="de">
```

Figure 4.39: Indication of language region targeted by Google: de = Deutschland
<http://www.garmisch-partenkirchen.de/en/db41673c-3301-78c8-8a27-52cc9b4d45f6.html>

Where the “de” for German is replaced in the tab by “en” for English the web spider will become aware that the site contains content in English and will display search results linked to that content. However, even if the language filter were correct and the search result linked to the English homepage, the description would still be incapable of functioning cohesively as it is not recognised by TL user and cannot be recovered through reference to the search query.

4.2 Conclusion

Chapter IV presented the findings from the analysis of Online Search and discussed the implications where the issue of cohesion is not adequately addressed in the translated web site. It was demonstrated that cohesion in Online Search is dependent on defining keywords that reflect not only the content of a web page but also the user language that will be used when searching online for such content. It was further shown that cohesion is dependent on the strategic placement of these keywords both on the visible web site and in the relevant HTML text, so that they can create a cohesive link, through repetition, between web site, SE and search query. Finally, it was shown that cohesion in Online Search contributes to the overall Google ranking of a web page, which, therefore, has a bearing on the ranking

of the web site in the list of search results and consequently on the general visibility of and access to the web site.

Recommendations on some of the issues outlined in this chapter are included in the final chapter of this dissertation. A supporting checklist can be found in Appendix B.

Chapter V presents the key findings in terms of Navigation that are pertinent to cohesion and the translated web site resulting from the analysis of the corpus web sites. The implications for the translator are outlined.

Chapter V

Analysis of Cohesion in Navigation

Chapter V

Analysis of Cohesion in Navigation

Introduction

This chapter presents findings from the analysis of navigation systems on the English corpus web sites. Three topics in particular are relevant in the context of the cohesion of a web site: general access to target-language content, the language of link labels and their destination content and the use of site-specific search features. The implications for the translator are discussed.

5.1 Navigation systems

The overall navigation system of a web site is fundamental to web site cohesion and therefore to a good user experience. Access to and successful navigation of TL content on a web site is dependent on the cohesive devices of repetition and superordinate/hyponym, in addition to non-verbal devices such as flag icons as TL indicators. Cohesion enables the user to move through the ‘hypertextual trajectory’ with ease, providing a sense of location and context and an immediate understanding of ‘where to go next’. This sense of orientation is all the more important in an online environment in view of the fact that the user can access a web site on any page of a site and not simply via the homepage.

Three findings in particular are significant in the context of cohesion and navigation

in the translated web site:

- Access to TL content is problematic on many of the corpus sites due to the manner in which this content is signalled as part of the global navigation system of the sites. On some sites, content can be easily accessed in the language indicated. However this is not the case on many of the corpus sites.
- At times, the language of the link labels in the corpus web sites does not correspond to that of their destination content. For example, a SL label links to TL content or a TL label has SL destination content, both of which prevent the creation of a cohesive link between link label and destination content.
- Site-specific search features are becoming increasingly important on web sites. This is due firstly to the increasing levels of content on web sites. It is also due to user expectations, whereby the user no longer simply hopes to find specific content in the TL but expects to find it and does not wish to spend time negotiating the navigation system of the site in order to access this content. The results returned in site-specific searches vary greatly in format, clarity and overall as indicators of content relevant to the search query.

5.1.1 Indicators of TL content

All ten web sites in the corpus have content in one or more languages other than the source language, German. The number of languages available varies from site to site as does the level of TL content available in each of these languages. In general, the web sites of the larger cities such as Berlin and Munich tend to have more extensive

TL content than those of the smaller cities such as Stralsund or Regensburg. While the web sites have relatively comparable navigation systems in place to enable access to this content, the ease of access nonetheless varies widely from one site to the next.

The flag icon is a common global navigational means of indicating the presence of TL content and is used on most of the corpus sites. The flags are for the most part representative of specific countries and are intended to indicate the presence of content in the language or main language of that country. In the following example from the Frankfurt site the presence of content in English is indicated by the British flag. Arabic, however, is indicated by what appears to be the flag of the League of Arab States, as seen in Figure 5.1. (far right):



Figure 5.1: TL indicator, Frankfurt site

(www.frankfurt.de)

Some web sites use text to indicate foreign-language content while others use a combination of text and icon:



Figure 5.2: TL indicator, tourism page, Münster site

(<http://www.muenster.de/stadt/tourismus/index.html>)



Figure 5.3: TL indicator, Leipzig site

(<http://www.leipzig.de/>)

On most of the sites the icon group indicating the presence of foreign-language content is placed towards the top of the page, close to the horizontal navigation bar. As seen in Figure 5.2., TL content in the five languages English, Dutch, French, Spanish and Japanese is indicated through an icon list and also textually in each language on the German homepage of the Münster site. This icon/text indicator of TL content is, however, placed at the bottom left-hand side of the page. This may require the user to scroll, depending on the browser used and the size of the screen used. On all other pages of the Münster site reference is made in iconic and textual form to English-language content only. The one exception is the page entitled ‘Skulptur Projekte Münster 07’, which indicates the presence of content on the site in Dutch, French, Spanish and Japanese. English, however, is not mentioned here. On the Regensburg source web site the only indication that the web site contains content in nine additional languages is the icon bar placed on the homepage as seen in Figure 5.4 below:



Figure 5.4: TL indicator, German homepage, Regensburg site
 (<http://www.regensburg.de/>)

Regardless of which flag icon is clicked on this bar, the page automatically defaults to a page with text in German and English, see Figure 5.5:

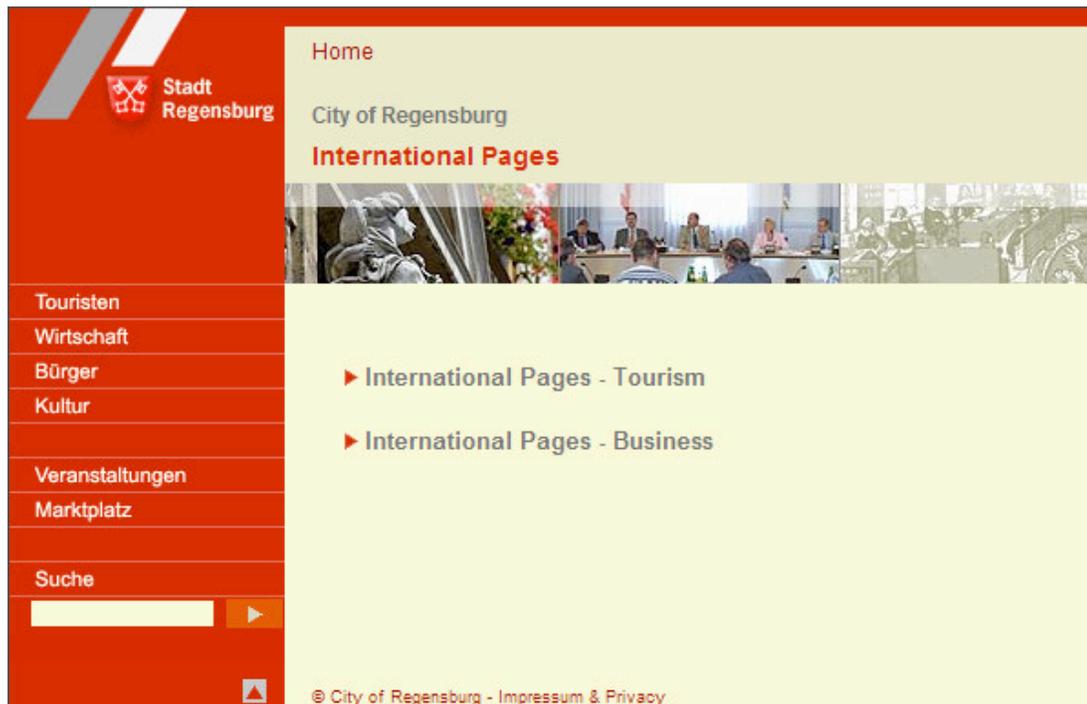


Figure 5.5: Destination page for all TL flag icons on German homepage, Regensburg site (<http://www.regensburg.de/international/index.shtml>)

As seen above, this page contains a navigation bar in German on the left-hand side of the page. The user has two main options or links on this page: a link to 'International pages – Business' and a second a link to 'International pages – Tourism'. If the business link is clicked, the user accesses a page offering two further options: a link to 'English business information' or a link with the label '[Informace pro investory a podnikatele](#)' to content in Czech. The navigation bar on the left-hand side of this page lists the languages 'English', 'Cesky' and 'German'. Where the 'Tourism' option is selected, the user will access a page that contains

some text in English in the centre, a link to the actual English content section of the site and a menu bar on the left and one on the right. The bar on the left contains links to content in ten languages, including English and German, indicated for the most part textually in each actual language. The navigation bar on the right-hand side contains nine flag icons with the word ‘Welcome’ translated into the languages represented by each national flag icon. A click on any of the tabs or icons left or right will allow the user to access content in the language indicated.

It would appear from the German homepage on the Garmisch site that the site only has TL content in English:



Figure 5.6: TL indicator, German homepage, Garmisch-Partenkirchen site
(<http://www.garmisch-partenkirchen.de/de/index.html?index=01>)

However, all pages of the ‘Tourismus’ section of the source web site have a drop-down menu labelled ‘International’ which has links to content in a further twelve languages. Extensive content is available in English only on this site. The content provided in the twelve languages other than German and English consists of a one-page piece of body text in the language indicated, but within a SL navigation system.

The other two main sections of the Garmisch source web site, ‘Bürgerservice’ and ‘Marktplatz’ make no mention of and provide no links to TL content.

The SL pages of the Berlin web site carry an icon of the British flag that would suggest, as was the case with the Garmisch-Partenkirchen site, that TL content on the web site is available in English only:



Figure 5.7: TL indicator, Berlin site
(<http://www.berlin.de/>)

However, the indicators of TL content as seen in Figure 5.8 are present on the page entitled ‘Städteverbindungen Berlins im Ausland’:

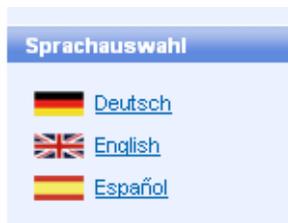


Figure 5.8: TL indicator, page ‘Städteverbindungen Berlins im Ausland’, Berlin site
(<http://www.berlin.de/rbmskzl/staedteverbindungen/index.html>)

The navigation bar in Figure 5.9 below is present as a global navigation feature on

all English-language pages of the Berlin web site:

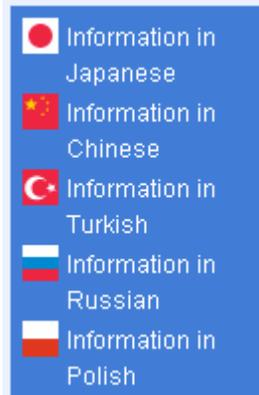


Figure 5.9: TL indicator on English pages of Berlin site
(<http://www.berlin.de/english/>)

A click on any one of the links in Figure 5.8. takes the user directly to content in the stated language.

Figure 5.10 is an example of a reference in English to an online information package for those who are new to Berlin and which is available in a number of languages.

The reference and link is found on almost all of the English-language pages on the Berlin site:



Welcome to Berlin!

Housing, working, child-care and more basic facts for new citizens. If you just moved to Berlin this is the information-package you need. Get your [free pdf-download](#) (1,5 MB). [more »](#)

Welcome to Berlin - other languages

The welcome information-package is also available in Arabian, French, Polish, Turkish, Russian, Serbo-Croatic and German. [more »](#)

Figure 5.10: English as SL for other languages
(<http://www.berlin.de/english/>)

Figure 5.11 advertises the same text in German:



Willkommen in Berlin!

Alles Wissenswerte für Berlin - "Anfänger" in verschiedenen Sprachen

Figure 5.11: TL content advertised in the SL
(<http://www.berlin.de/lb/intmig/>)

The German version is carried only on the pages of the section with the heading

,Der Beauftragte für Integration und Migration in Berlin’.

5.1.2 Issues with indicators of TL content

One of the most important aspects of the provision of multilingual content on a web site is to ensure that the user is aware of this content. Given that the user can access a web site through any one of its pages and not simply through the main or homepage, the indication of multilingual content must be treated as a significant global navigational tool which is fundamental to a cohesive web experience. As such, it is vital that this global navigation tool be truly ‘global’ in that it is provided consistently on all pages of the web site, regardless of the extent of the multilingual content or the number of languages in which content is available.

The Munich and Frankfurt web sites and, for the most part, the Cologne site, provide links to multilingual content on all of their web pages. However, the provision of a multilingual indicator on the source homepage only, as is the case with the Regensburg site, or the indication, falsely, of content in only one foreign language, as on the Berlin site, means that the user will remain unaware of the wealth of information provided in languages other than German and English on these sites.

The inconsistency seen on various pages of the Münster site, indicating content in five or four languages or just in one language, could prevent the user from accessing content in a language that is not indicated on a particular page, despite the fact that a considerable level of content may be available on the site in that language. The content available in various languages on a web site should, therefore, be flagged prominently on all pages of the site.

It is, however, not sufficient simply to flag or indicate this content on all pages, be it in textual or iconic format. The user must be guided directly to the content in the chosen language and should not be forced to first go through a series of clicks and pages, often in a language other than the one selected, before finally being able to access content in that language. Content on the Berlin site in a language other than German or English is accessed, for the most part, through the English content. It is perhaps assumed that the non-German speaker will automatically access content in English.

A similar issue exists on the Regensburg site where the user, having clicked the icon on the homepage for a language other than German or English, is forced to go through a number of pages that are in English and German in order to access content in one of the other languages offered. It must be assumed that if a user is looking for a specific language, s/he will only understand that language. Therefore, it is unlikely that s/he will persevere in the attempt to access content through an unfamiliar language. This could result in the user leaving the site and not returning. The TL content that has cost the company or organisation much time and money will not receive the traffic and business it deserves. The translator must attempt to provide a cohesive web site for the user and in so doing enhance the credibility of the site and the loyalty of the visitors to the site.

The use of flag icons to indicate TL content is the most common approach taken by the web sites in the corpus. However, the flag icon is not without its problems. One of the main issues with this type of language indicator is its size. As these icons are

frequently quite small, it is difficult for the user to establish which country the icon represents and, therefore, which language it symbolizes. This requires the user to click and access the destination content in order to see which language the content is in. In addition, though the icons are generally placed in a position at the top of the web page where they can be easily seen, they are often surrounded by advertising or other links and information and are not sufficiently prominent. Given the importance of salience and framing in terms of cohesion on a web site, the TL indicators should be clearly visible, regardless of whether they are in iconic or textual form.

A further issue regarding the use of flag icons and one that is not insignificant is the fact that each icon represents a country and not specifically a language. One exception here is the use of the flag icon of the League of Arab States on the Frankfurt site which is used to indicate content in Arabic. Nielsen (1996) recommends indicating the presence of content in up to seven languages by listing the word for each language in that language. He suggests that a list of eight or more languages would be difficult to scan and should be in iconic form. However, given that the user will automatically scan for his/her own language, it seems reasonable to suggest that it is easier to scan a clear and precise list of languages in text form than a set of flag icons that are quite small and not very clear. The use of text to refer to different languages as opposed to national flags is, in addition, more 'inclusive' and politically neutral. The word 'Deutsch' for example, would appeal to all countries and regions where German is the official language and not, as is the case with the German flag, simply to Germany.

5.1.3 The language of link labels and their destination content

This section examines an issue that is crucial to navigation on a multilingual web site and consequently to cohesion on the site. It involves the language of link labels and that of their corresponding destination content.

5.1.3.1 SL link labels as part of TL content

There are examples in the corpus web sites where the global and the local navigation systems have link labels that are in the SL despite being part of the foreign-language content of the site. This is the case with the Berlin and Garmisch-Partenkirchen web sites. The English pages of the Berlin web site contain a navigation bar that is horizontal and is placed towards the top of the page. It is identical to the navigation bar on the German SL pages (see below). The English pages have an additional navigation element, also present on the SL pages, which consists of a group of five links in the SL, placed above the primary navigation bar, that include ‘Stadtplan’ and ‘Branchen-Info’:

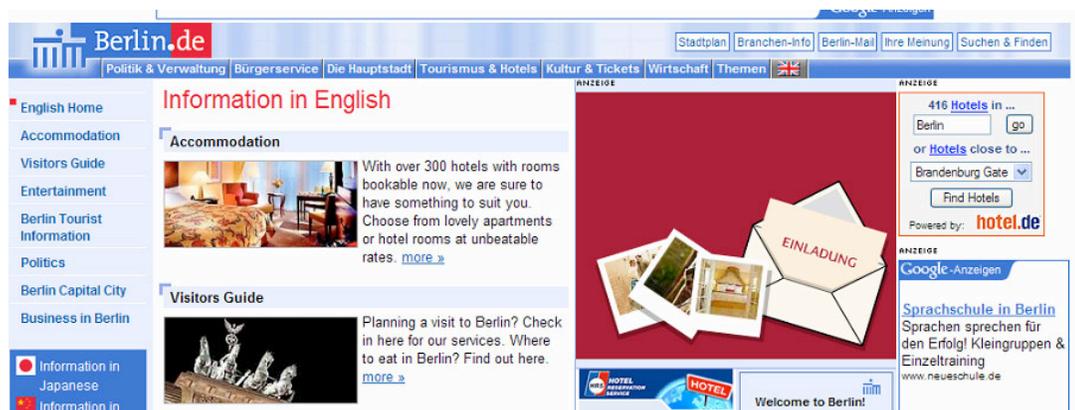


Figure 5.12: Search feature ‘Suchen & Finden’, Berlin site (<http://www.berlin.de/english/>)

The primary navigation bar and the set of five links above the navigation bar are in the SL on all English content pages of the Berlin site. The actual navigation tabs for the English content are placed vertically on the left-hand side of the page and include the tabs ‘Accommodation’ and ‘Visitors Guide’ (see Figure 5.12).

Unlike the Berlin site, the main navigation bar on the English content pages on the Garmisch-Partenkirchen site, which is placed horizontally towards the top of the page, has been translated. However, there is a second set of tabs placed vertically on the left-hand side of each page that also serves as a global navigation tool. These tabs which include labels such as ‘Suchen und Buchen’ and ‘Gewinnspiel’ are in the SL on all TL content pages:



Figure 5.13: Tabs in the SL, left, Garmisch-Partenkirchen site
(<http://www.garmisch-partenkirchen.de/en/db41673c-3301-78c8-8a27-52cc9b4d45f6.html>)

The global navigation bars and secondary links that are in the SL on the English pages of the Berlin and Garmisch sites link to SL destination content.

A list of ‘Useful links’ is provided on the Stralsund web site on the page entitled ‘Business / Stralsund Network for Business Start-Ups’:



Figure 5.14: List of ‘Useful links’, Stralsund site
(http://www.stralsund.de/hst01/content.nsf/docname/Webseite_6124FBC1071FFC00C12570DD002B9009?OpenDocument)

The web sites listed above provide information on various issues involved in setting up a business. The link labels for each web site as shown here consist of the domain name for that site. With the exception of www.enterprise-mv.de and www.change-online.de, the labels are in the SL. The destination content of each of these links is also SL. However, three of the web sites listed have English content.

The following three links in the SL are on the main TL page for the ‘City Hall’ section on the Frankfurt site. They have SL destination content, which is noted in brackets:



Figure 5.15: SL links as part of TL content, Frankfurt site
(http://www.frankfurt.de/sixcms/detail.php?id=stadtfrankfurt_eval01.c.123086.en)

Reference is made to the destination content of these links in the body text of the City Hall page. However, the SL link label (e.g. Amt für multikulturelle Angelegenheiten) is buried in the middle of a long paragraph of text, as in Figure 5.16:

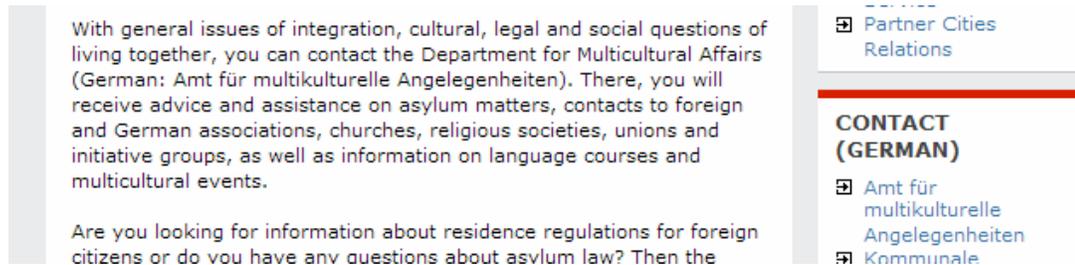


Figure 5.16: Reference to SL links in body text, Frankfurt site
(http://www.frankfurt.de/sixcms/detail.php?id=stadtfrankfurt_eval01.c.123086.en)

The SL text is in the middle of the TL body text and is not highlighted in any way. There is no obvious link between it and the links on the right of the page, which is in the ‘New’ position and therefore less likely to be consciously read.

Figure 5.17 is from the main page of the ‘Tourism’ section of the Frankfurt web site:

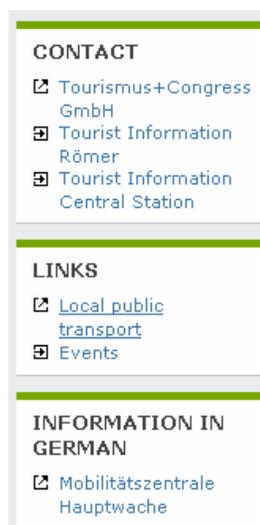


Figure 5.17: SL links in TL content, Frankfurt site
(http://www.frankfurt.de/sixcms/detail.php?id=stadtfrankfurt_eval01.c.124836.en)

The first link ‘Tourismus+Congress GmbH’ links to the English content of an external web site that has much information that could be of use to the tourist in Frankfurt. The last link in the list, ‘Mobilitätszentrale Hauptwache’, links to content in the SL, and is similar to the content accessed through the link ‘Local public transport’. Figure 5.18 is an example of how English has become the SL for languages other than German on the Berlin site:

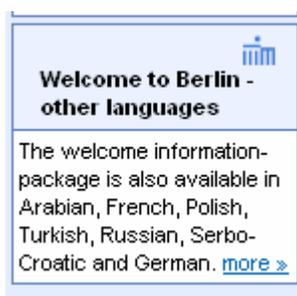


Figure 5.18: English as SL on the Berlin site
(<http://www.berlin.de/english/>)

The link ‘Welcome to Berlin – other languages’ links to a page with SL text followed by a list of links to a document, available in a number of languages, with the title ‘Willkommen in Berlin’. This is a booklet written by the Commission for Immigration and Migration that provides advice on many aspects of life in Berlin. The document is not available in Serbo-Croatian (sic), as mentioned above, but is available in Spanish which is not included in the list of languages.

5.1.3.2 Issues with SL link labels as part of TL content

The inclusion of SL link labels as part of the TL content of a multilingual web site can seriously affect cohesion on the site and consequently the user experience. This is particularly critical where the global navigation system contains SL text. The

global navigation system, as previously mentioned, is important for orientation particularly when the user accesses a web site on some obscure page, as it provides the user with information on the main sections of a web site and enables him/her to go to any of these sections directly or to the homepage. However, if the labels of the navigation bar cannot be understood by the user, the navigation bar can no longer function as the powerful cohesive device it should be.

In the case of the German navigation tabs on the English pages of the Berlin site, force of habit could result in the user clicking one of these tabs, despite not understanding the text, given that the bar is in the most common position on the page for navigation bars. A quick scan of the page could also result in the English-speaking user confusing the tab label 'Tourismus' with the English word 'Tourism' and clicking the tab as it is recognisably a link to one of the main sections of the web site. Once clicked, the tab will take the user to SL content and a confusing and irritating Web experience.

This problem is compounded by the actual English-language navigation bar on the site. As mentioned in Chapter II, visual disconnectedness can aid cohesion on a web site in that it sets the items in question apart from others on the page and, as a result, aids scanning. The individual tabs on the English navigation bar on the left of the page, however, are lacking in any type of framing element that would set them apart from each other and indicate that they are navigation tabs and not simply links. As a

result, the user is even more likely to click the horizontal tabs on the Berlin site because they are 'visually disconnected' from each other but cohesive for the purposes of the web site. If the navigation tabs cannot be translated and linked to the English-language content on the site, they could be removed from the English-content pages and the English tabs proper could be optimized to look like navigation tabs.

The issue of SL tabs in the navigation bar on the Garmisch site is somewhat less critical given that the main navigation bar is provided in English. However, the inclusion on each English content page of SL links in a global navigation tool could lead to some confusion. As destination content for some of the links in this bar is available in English, the tabs could be translated and linked to this content. Those links which have no corresponding content on the site in English could be removed from the English content pages.

The provision of a list of 'useful links' to web sites that are related in some way to the current content is a common feature of web sites today. Such a list helps on the one hand to create a link network of thematically-related and high-value web sites in the sense of SEO. It also offers the user added value by suggesting other relevant web sites. If the content turns out to be as indicated and is useful, the user will most likely return to the original web site as it has gained a certain level of credibility through its reliability.

The list of 'useful links' as part of the English content of the Stralsund web site as shown and discussed above is unlikely to be of value to the non-German speaker as

the labels consist of the domain names for the sites, most of which are in German. The links are to SL content despite the fact that three of the sites have content in English: www.bmwa-bund.de, www.existenzgruender.de and www.kfw-mittelsstandsbank.de. The site www.change-online.de does not appear to be live at the time of access.

Overall, it is preferable to provide a descriptive link label when providing a link, be it to another web site or another page on the same site. On a multilingual web site the label should also be in the appropriate language. Where the destination web sites have English content, as is the case with the three sites mentioned above, the link should be directly to the English content and not simply to the homepage of that site. If web sites that do not have content in English have to be included in the list, a note should be added informing the user that the destination content is available in German only. This will allow the user to make an informed decision as to whether the information could be useful and should be accessed despite the fact that it is not in English.

This also applies to the example of the three links on the Frankfurt site, ‘Amt für multikulturelle Angelegenheiten’, ‘Kommunale Ausländer- und Ausländerinnenvertretung’ and ‘Internationale Angelegenheiten’, as shown in Figure 5.15 above. While it may not be possible or desirable to have the entire content of a web site translated, if a web site contains content that might reasonably be considered to be of particular interest to the TL user, then the user should at least be provided with a reference to the destination content. One possible solution is to

translate the link labels into the target language and include a note that the destination content is available only in German, e.g. 'Contact (in German)'. In this way, the user can make a decision as to whether s/he wishes to have the text translated. Equally, the user may simply require the contact details of the organisation listed on the destination page.

Where a user is in the process of accessing specific content by working his/her way through a series of TL links, a single link that is in the SL will stop this cohesive path, as it is unexpected, cannot be understood and will not enable the user to presuppose the content of the destination page of that SL link. A link label that cannot be understood, for whatever reason, is redundant. Relevant content in the correct language may only be a click away but the user will remain unaware of it due to the lack of cohesion resulting from the barrier created by a label in German.

The issue involving the document 'Willkommen in Berlin/Welcome to Berlin' is a case in point. As mentioned, this document provides detailed and useful information in seven languages on many aspects of life in Berlin. The target group includes people from different linguistic and ethnic backgrounds who are relocating to Berlin or who simply need specific information about Berlin. The fact that the document is translated into seven languages is a clear indication of the importance attached to it by the commissioning authorities. However, as the presence of this document appears to be flagged only in German and English, and not in the actual languages in which it is available, it is highly probable that the document will not be found or accessed by the very people for whom it was originally written. It is not clear

whether this document can be accessed directly through the individual pages of the other target languages, for example through the Turkish and Arabic pages.

The use of SL link labels as part of TL content could have a negative effect on the SEO work that has been carried out on a web site, thus affecting its overall ranking with search engines and its listing in a SERP. To begin with, the link label may normally be a keyword for the destination page. As it is in the SL, the cohesive value of the term as a keyword link will be lost. In addition, a search engine will most likely fail to match SL link labels with the keywords placed in the page title, page heading or body text of the destination page because these are in the TL.

As a rule, the inclusion of SL labels as part of TL content should be avoided, as it serves only to confuse the user and consequently affect the sense of orientation and context that is vital in order to successfully negotiate a web site. Content that may be valuable to the user will be withheld and overall, the credibility of a web site could be affected.

5.1.3.3 Link labels in the TL with SL destination content

Link labels in English that link to German destination content are, in the majority of cases, elements of the local navigation system of a web site. The destination content can be a document, such as a PDF file or an internal or external web page.

The following English menu tabs ‘Concerts’ and ‘Tickets’ are part of the ‘What’s

on' main section on the Cologne site:

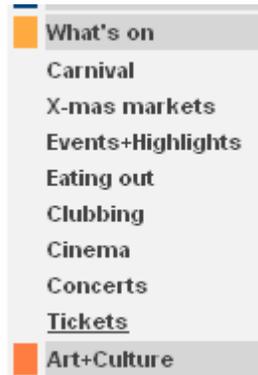


Figure 5.19: TL link with SL destination, Cologne site
(<http://www.koeln.de/en/whatson/whatson.html>)

The 'Concerts' tab links to a SL page on the same site entitled 'Konzerte in Köln', which is a guide to all concerts and events taking place in and around the city of Cologne. The 'Tickets' tab links to an external online ticket shop called 'Köln Ticket' that does not have English language content.

The first two links in Figure 5.20 from the Stralsund web site link to PDF files from the Pommersche Volksbank eG web site while the third link is not currently active:



Figure 5.20: TL links to PDF files in SL, Stralsund site
(http://www.stralsund.de/hst01/content.nsf/docname/Webseite_6124FBC1071FFC00C12570DD002B9009?OpenDocument)

Both PDF files are in German and an English version is not available.

The information box in Figure 5.21 from the Munich site offers two links leading to more detailed information on the weather: 'Detailed forecast' and 'more weather?'

wetter.com':

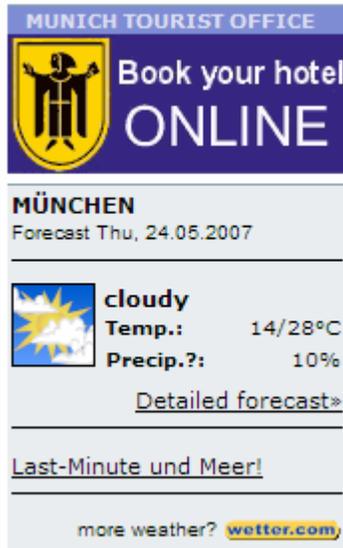


Figure 5.21: TL links with SL destination content, Munich site
(<http://www.muenchen.de/home/60093/Homepage.html>)

The destination web page and site are identical for both links, are in German and do not have content in English.

The Munich site has information on the Auer Dult, a famous market that takes place there each year. The user can access information about this market in the English content pages through the quick link 'markets in Munich' on the left-hand side of each page. In this case the destination content is in English and is part of the www.muenchen.de site. However, there is also a link on the 'events' page of the 'Tourism' section labelled 'further information on the Auer Dult' which links directly to the official Auer Dult web site. It links to the German content on the site, despite the fact that the site has content in English.

5.1.3.4 Issues with TL link labels with SL destination content

The expectation of the user when clicking a link in a particular language is that the destination content will also be in that language. If the destination content is in any other language it will cause confusion and frustration to the user and will most likely lead him/her to either press the 'Back' button and return to the previous page or leave the site altogether. There are generally two reasons why a TL link will link to SL content. The first is that the information is simply not available in English. The link label has been translated and left in place, that is to say, linking to SL content. The second reason is that the address that has been inserted into the link is not the specific address for the relevant English content but is perhaps the German homepage address. This is the case with the link from the Munich site to the Auer Dult site which is www.auerdult.de. The correct address which should be inserted into the link and which would take the user directly to the English-language content is: <http://www.auerdult.de/englisch.html>. While there is a link to English-language content on the German Auer Dult homepage, it is at the bottom of the page and not immediately visible unless the user scrolls.

In order to ensure cohesion, it is important that the translator does not simply translate the link label but that s/he also ensures that the destination content is as indicated by the link, that is to say that the content is correct and that the content is in the target language. If content is available in English, every effort should be made to link the user directly to that content.

As seen with the example from the Cologne web site, the provision of information on different events including concerts, theatre and cinema is a problem for many

web sites in the corpus. This arises from the nature of an events calendar which is updated regularly and, therefore, would require ongoing translation in order to remain up to date in the TL versions. Many of the corpus sites provide a link to information on events in German and include a note to the effect that it is only available in German. However, the Cologne site allows the TL user to link to German content without prior notice that the destination content is available only in German. The Frankfurt web site has made an attempt to deal with the issue by providing a link which allows the user subscribe to a weekly e-newsletter providing up-to-date information, in English for the most part, on events in the city. A further solution would be to simply link the events tab to an external site, for example in the case of Cologne this could be the City Guide at <http://www.wcities.com/en/guide/entertainment/32/guide.html> which provides a guide in English to theatres, night clubs, restaurants and cinemas.

A TL link should, as a general rule, not be placed on a page unless there is content in the same language to match it, be it on the same site or on an external site. In exceptional circumstances, a link could be provided to content that is deemed to be of great importance to the user and where the user is informed that the destination content is in the SL. In this way, the user can make the final decision as to whether to access the content in German or not and perhaps have it translated.

5.1.4 Site-specific search features

A site-specific search feature is considered to be part of the supplemental navigation system of a web site. A search feature effectively puts the user ‘in the driver’s seat’ and also allows for a high level of specificity (Rosenfeld and Morville 2002 p127).

Nielsen (2001) suggests that a site-specific search feature can be seen as an ‘escape hatch’ when the user cannot find a reasonable place to go next.

Six of the ten web sites in the corpus have a search feature for both the German and the English content: Munich, Berlin, Frankfurt, Leipzig, Regensburg and Cologne. Hamburg has a search feature for the German content but not for the English. The other three web sites do not offer site-specific search.

The search features consist of a box into which the user types a search query and a button which is clicked to begin the search. The label for the search button on five of the sites, with the exception of Berlin, reads ‘Search’. On the TL page of the Berlin site the user has to first click a link labelled ‘Suchen und Finden’ before accessing the page with the search box. The Cologne search box is on the left-hand side of the page, as is the search box on the Regensburg site. Both would appear to be visible even on the smallest screen, without requiring the user to scroll. The search features on the other four sites are situated on the right-hand side towards the top of the page, in a prominent position.

The analysis of the search features on the six sites was carried out using the single query term ‘museums’, reflecting information that is available on all the sites and that is likely to be searched for by the English-speaking visitor to the cities.

5.1.4.1 Site-specific search features - search results

As mentioned above, the label for the link leading to the search feature on the English pages of the Berlin site is in the SL. When it is clicked to access the search feature proper, it defaults to a German page:



Figure 5.22: SL search label in TL content, Berlin site
(<http://www.berlin.de/themensuche/.bin/index.php>)

Despite the fact that the link label is in the SL and would most likely not be understood by the TL user, a search was carried out for content on ‘museums’. The search returned a long list of results, many of which appear to be relevant and useful and are presented under two headings, the first being ‘Angebote’ and the second ‘Ergebnisse der Volltextsuche’. The first two results in the ‘Angebote’ list are in English and provide links to an extensive list of museums in Berlin:

[Museums and Art Galleries](#)

pergamonmuseum Berlin has more than 170 museums and collections. They offer everything from the ancient to the ultra-modern - the Egyptian Nefertiti, the greek Pergamon Altar, the painters of the " ...

■ [Museums and Art Galleries](#)

Berlin museums and collections. More than 170 institutions offer everything from the ancient to the ultra-modern.

Figure 5.23: Results of site-specific search for ‘museums’, Berlin site
(<http://www.berlin.de/themensuche/.bin/suche.php?x=museums&db=site>)

The remaining results are either English or German.

The same search on the Munich site returned many pages of results with ten results on each page. The first three results are:



Figure 5.24: Results of site-specific search for ‘museums’, Munich site
(http://www.muenchen.de/home/81049/search_page.html?search_string=museums)

The first result is in English and links directly to a page in English about museums in Munich. The remaining results on the first page are in a number of languages including Arabic, Russian and German.

When carrying out a search on the English pages of the Cologne site, the results

automatically default to a SL page:

Suche

Geben Sie einen oder mehrere Suchbegriffe ein und wählen aus, wo Sie suchen wollen.

museums Alle Suche

Empfehlungen der Redaktion:

- [Übersicht über die Kölner Museen](#)

254 Treffer auf www.koeln.de:

- [Cologne - Museums](#) 19.10.2006
- [Museums \(KölnTourismus\)](#) 02.06.2007
- [Museums \(KölnTourismus\)](#) 07.06.2007
- [Museums \(KölnTourismus\)](#) 02.06.2007
- [Museums \(KölnTourismus\)](#) 02.06.2007

[... mehr/ausführlichere Treffer](#)

Figure 5.25: Results of site-specific search for ‘museums’, Cologne site
(<http://www.koeln.de/suche/suche.php>)

With the exception of a group of links that are partly in English (Cologne, Museums), and link to English-language content, all other elements of text on the page are in the SL. The first result in English, ‘Cologne – Museums’, links to relevant content in English on the Cologne site. The four links entitled ‘Museums (KölnTourismus)’ have identical destination pages in English but have different page headings. The first and fourth have the page heading ‘Attractions’ while the other two are ‘Guided Tours’ and ‘Welcome Card’ respectively.

A search on the Frankfurt site returns ten results, all in English The first result links

to a page entitled 'Culture':

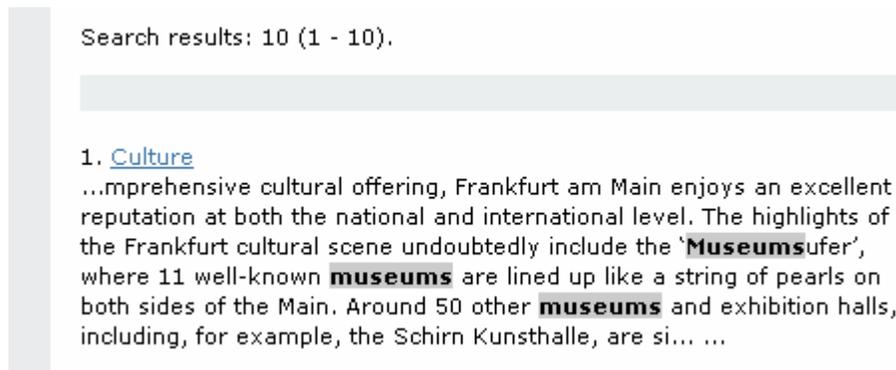


Figure 5.26: Results of site-specific search for 'museums', Frankfurt site

(http://www.frankfurt.de/sixcms/detail.php?id=stadtfrankfurt_eval01.c.123086.en&_ffmpar_az%5B_gsid_inhalt%5D=stadtfrankfurt_eval01.c.1806772.en&_sprache=en&_searchval%5B_vtSearchmode%5D=fulltext_all&_searchval%5B_vtSuchwoerter%5D=museums)

This is the entry page for the main section that provides information on museums, historical buildings and theatres, among other things.

5.1.5 Issues with site-specific search features

The provision of a successful site-specific search feature is a complex and challenging matter and one that, in its entirety, goes beyond the scope of this study.

However, those issues that relate to cohesion and translation are addressed here.

The first concerns access to the search feature on the English content pages of the Berlin site. As the link label is in the SL, it will not be understood by the user and, given that it is a link and not a search query entry box, it will not be recognised as being a link to a search feature.

Secondly, the manner in which the search results are presented on the Cologne web site lacks cohesion and could cause confusion for the user. Although the list of results is a mixture of English and German, on first sight the user could be led to

believe that the entire page is in the SL. Text that presents a combination of languages cannot function cohesively.

With the exception of the Cologne results, those returned on the other web sites are presented in a manner similar to full Web searches carried out using a search engine such as Google or Yahoo!. As is the case with such a Web search, the information returned in a site-specific search result uses the text defined in the title tag and the description content metatag for the page in question. Issues with site-specific search regarding the optimization of tags and metatags in terms of SEO are similar to those discussed in Chapter IV.

5.2 Conclusion

Chapter V presented the findings of the analysis of the corpus web sites in terms of Navigation. It was found that access to TL content is not always straightforward, often for reasons of lack of cohesion. It was shown that cohesion on a multilingual web site depends on a link being created between lexical items in the same language. The increasingly important navigation feature, site-specific search, was seen to return results with varying levels of usefulness. As with search using one of the major search engines, a cohesive link between the relevant web site content and the search query cannot be created if the language does not match or if the search result produces a link that does not contain the keywords used by the user in the search query. The implications of the failure to address the issue of cohesion in navigation for the translated web site were discussed.

Recommendations on some of the issues outlined in Chapter V are provided in the concluding chapter to this dissertation. A supporting checklist can also be found in Appendix B.

Chapter VI will present findings from the third key area of web site cohesion: Page Content.

Chapter VI

Analysis of Cohesion in Page Content

Chapter VI

Analysis of Cohesion in Page Content

6.0 Introduction

Chapter VI presents the findings from the analysis of Page Content of the corpus web sites. Two elements of page content, the teaser text and widgets, are addressed and the implications for the translator are discussed.

6.1 Page content

Page content refers to all on-page textual features of a web site that are not classed as being part of the embedded and supplemental navigation systems of a web page and site, as treated in Chapter V. Page content, therefore, includes features such as teaser text, widgets, information boxes and continuous text. As previously mentioned, continuous text does not form part of this study as it can be analysed using traditional models of textual cohesion. This chapter addresses cohesion in teaser texts and widgets on the translated web site.

6.1.1 Teaser text

The significance of the teaser text as a cohesive device lies in the fact that it provides the user, scanning or skimming a web page, with brief information on additional content that can be accessed by clicking on the teaser text. This enables the user to make a quick decision, based on the information provided in the teaser text, as to whether the destination content of a teaser text is relevant to his/her requirements. In

teaser texts a cohesive link is created between the lexis of the teaser heading, the body text of the teaser and the destination content of the teaser text. Saliency, framing and information value play a vital role in cohesion. In order to function cohesively the teaser text must be concise, an accurate indication of the destination content and in the correct language. Where a number of teaser texts are grouped together in close proximity, each one should be clearly differentiated from its neighbours in terms of the destination content which it describes and to which it links.

The model for the analysis of cohesion in Page Content was tested on the teaser texts on the English homepage and on tourist-related sections of nine of the ten web sites in the research corpus. The Stralsund web site does not make use of teaser texts. There are two main findings in relation to cohesion and the teaser text which are significant in the context of web site translation:

- Teaser text headings are not capable of functioning cohesively in the TT as they are frequently untranslated, ambiguous or obscure. In addition, they do not always reflect the actual destination content that they are meant to describe and are therefore not capable of creating a cohesive link through lexical cohesion.
- Direct address and superfluous text are common features of teaser text body text on some of the English web sites in the corpus, which lengthens the TT and at times places the main keywords at the end of the text.

6.1.1.1 Teaser text headings are in the SL, are ambiguous or obscure

As seen in Figure 6.1 below, the teaser text heading, ‘Fressgass’, from the Frankfurt web site has not been translated into English:



Figure 6.1: Teaser text heading, Frankfurt site
(<http://www.frankfurt.de/sixcms/detail.php?id=317586>)

‘Fressgass’ is the colloquial name for this pedestrian zone in Frankfurt which is known for its wide variety of restaurants and cafés. The label ‘Fressgass’ is bolded and is a link and is separated from the following text by a space. The name, Fressgass’, is mentioned a second time in the first line of the body text of the teaser, which theoretically creates a cohesive link.

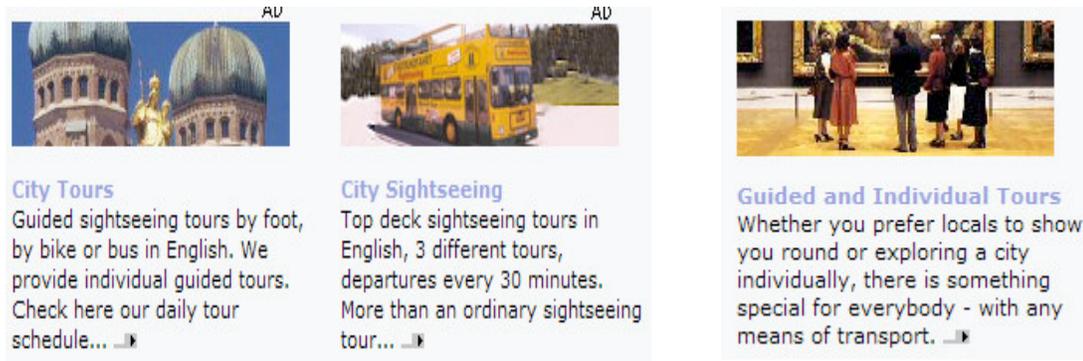
The teaser text heading in Figure 6.2, ‘What to do?’ on the Berlin site links to a page with further teasers on top events, museums, shopping and films in English:



Figure 6.2: Teaser text heading, Berlin site
(<http://www.berlin.de/english/visitors/index.html>)

The teaser heading is bolded and is a hyperlink.

The following three teaser texts are listed together on the same page in the ‘Tourism’ section of the Munich web site:



City Tours
Guided sightseeing tours by foot, by bike or bus in English. We provide individual guided tours. Check here our daily tour schedule... ➤

City Sightseeing
Top deck sightseeing tours in English, 3 different tours, departures every 30 minutes. More than an ordinary sightseeing tour... ➤

Guided and Individual Tours
Whether you prefer locals to show you round or exploring a city individually, there is something special for everybody - with any means of transport. ➤

Figure 6.3: Teaser text headings, on same page, Munich site
(<http://www.muenchen.de/Tourismus/Sightseeing/7565/index.htm>)

The three headings are set apart from the body text of the teaser through their colour and are all hyperlinks. The Web user could reasonably assume, based on the three headings ‘City Tours’, ‘City Sightseeing’ and ‘Guided and Individual Tours’, that the tours on offer are very similar if not identical in content, particularly as they all refer to tours of Munich.

6.1.1.2 Issues with teaser text headings in the SL, ambiguous or obscure

The teaser text headings mentioned above are ideal for scanning as they are prominent due to the fact that they are in bold and are highlighted in comparison with the body text. Despite the fact that there is theoretically a cohesive link between the teaser heading in Figure 6.1 above and its body text through the term ‘Fressgass’, such a link is in reality of little value in the TT as the word ‘Fressgass’ does not have any referential meaning for the non-German speaker. It is unlikely to be used in

online search by the English-speaking user. Therefore, a cohesive link between search query, search result and web page will be impossible.

The information in this teaser text is important for any tourist given the difficulty that people encounter in finding suitable restaurants or cafés in an unfamiliar city.

It is possible that the heading shown in Figure 6.2 above is meant to read 'Things to do', which would be somewhat more meaningful in the context. Nonetheless, a teaser header should be sufficiently descriptive to give an indication of the destination content or the subject area within which it functions, which means it should contain at least one if not more keywords relating to the destination content. This is not the case with this header. As the teaser text header contains no keywords, it cannot create a cohesive link through repetition either with the body text of the teaser or with the destination content of the teaser.

Any piece of text on a web page which is included in a SERP could be used by a SE such as Google in a search result. This teaser heading, if taken out of context in this manner, would be meaningless. This teaser text header is a hyperlink and is bolded and is therefore important for the web spider. However, the placement of a keyword such as 'Berlin' in the header would enable the web spider to properly index the page and return it in an online search. This would contribute to the overall ranking of the web page. Any change to this teaser heading would require the heading on the teaser text destination page to be revised and also the menu tab on the left-hand side of the screen labelled 'What to do?'. This would facilitate cohesion and provide additional valuable keyword links.

One of the most important requirements of a teaser text that is part of a group of teaser texts situated in close proximity to each other on a web page is that the heading will be salient and be clearly differentiated from its neighbours. The three teaser texts that link to information on sightseeing tours of Munich demonstrate the difficulty of differentiating between teaser headings, especially where the destination content is relatively similar. In a conventional text, the example of these three 'sightseeing' teaser texts would be considered to be strongly cohesive as a result of the lexical chains through which they are linked to each other, for example 'City tours', 'Guided sightseeing tours', 'City Sightseeing', '3 different tours' and 'Guided city tours and sightseeing tours'. However, it is reasonable to assume that the user will scan the texts, that is to say the headings, in order to establish which of the three, if any, is most relevant to the user requirements. The user may then choose to click one, if s/he can differentiate between the destination content of the three teasers. From an English-language point of view there is no apparent difference between the terms 'City tours', 'Sightseeing tours' and 'Guided and Individual Tours', given that they all refer to tours in and around the city of Munich. However, each of the teaser texts links to content about different types of tours of Munich. 'City Tours' offers a 'Ghost' tour and a 'Food tasting tour' at the Viktualienmarkt, among others, while 'City Sightseeing' offers tours of Munich on a double-decker bus, a 'Hop-on, hop-off' system. In each of these cases, the teaser links to an external web site. The third teaser 'Guided and Individual Tours' links to a page on the same site that contains a further three teaser texts labelled 'Guided Tours',

‘Special Tours’ and ‘Individual Tours’. While the text in these teaser headings appear to create a cohesive link with the body text and the destination content, in practice a clearer use of superordinate/hyponym would be of greater value for cohesion on the site.

6.1.1.3 Teaser headings do not correspond to the destination content

The phrase ‘Eating Out’, which is the heading in the following teaser text from the Cologne web site, has a specific meaning in English and suggests a restaurant guide or tips of some sort, as in ‘where to eat out in Cologne’:



Figure 6.4: Teaser headings and body text, Cologne site
(<http://www.koeln.de/en/index.html>)

The immediate content seen by the user having clicked this teaser text is a list of descriptions of various dishes that are native to Cologne and that have unusual

names:

■ Eating out

Cologne has not only a special beer to offer, there are also some meals that bear extraordinary names. For example the "Halver Hahn", which means "one half of a cock", is not really chicken. It is simple bread with cheese. Here you'll find some especialities:



"Himmel un Ääd"

means "heaven and earth" but if you order this, you'll get mashed potatoes with apple mash, blood sausage, onions and sometimes bacon.

Figure 6.5: Destination content of teaser text in Figure 6.4, Cologne site
(<http://www.koeln.de/en/whatson/eat.html>)

A link to restaurants that serve the type of food described here is provided at the bottom of the page ('In these restaurants ...'), which will only be seen if the user scrolls to the bottom of the page. The label text for this link is part German (kölsche Küche), contains a term that is often seen in 'German-English' today in a culinary context (Gastroguide) and the destination content is in German, as indicated in the brackets:

In these restaurants you'll get the "kölsche Küche":
"Gastroguide at koeln.de" (German)

Figure 6.6: Link to restaurants implicit in Figure 6.4., Cologne site
(<http://www.koeln.de/en/whatson/eat.html>)

The following teaser from the Leipzig site is a further example of a heading that

indicates one thing but delivers another:

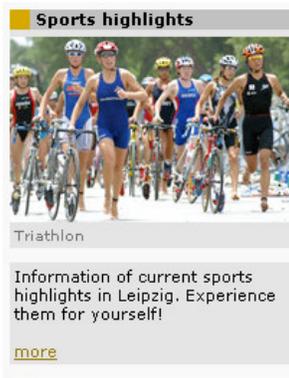


Figure 6.7: Teaser heading does not correspond to destination text, Leipzig site
(<http://www.leipzig.de/int/en/>)

This teaser entitled ‘Sports highlights’ links to a page that has two further links with the accompanying text: ‘Events organised by Leipzig’s sports associations are listed under:’ and ‘A list of current events in Leipzig is accessible at:’ respectively:



Figure 6.8: Destination content of Figure 6.7, Leipzig site
(<http://www.leipzig.de/int/en/sport/aktuelles/>)

The first link in Figure 6.8 is to the Leipzig Sports Association web site which has content only in German. The second link is to a page of events in Leipzig that includes not only sports events but also concert and theatre listings.

6.1.1.4 Issues with teaser text headings that do not correspond to their destination content

The teaser heading 'Eating out' on the Cologne web site is an example firstly, of creating expectations that are not fulfilled and secondly of the difficulty of defining a concise and accurate heading in the case of destination content that consists of different types of information.

The expectation created in the teaser heading is that the destination content will be a restaurant guide for Cologne. However, the destination content is somewhat spatially incohesive in that the information promised in this heading does not immediately materialise when the teaser is clicked, as it is situated at the bottom of the destination page, where it will be seen only if the user scrolls. The body text of the teaser does make reference to the information that 'Cologne has ... meals that bear extraordinary names' (sic), which is the content that is immediately visible on accessing the destination page. However, in view of the Eyetrack III study (Outing and Rual 2004) which suggests that where there is a large disconnect or gap between a teaser heading and its accompanying text, the user may not read past the heading, it is quite possible that in this case the user will decide not to read beyond the 'gap', which in this case is created by the image of a meal. As a result, the information provided in the teaser heading is all the more important.

This label for the link to the list of restaurants is partly in German and its destination content is also German. A cohesive link cannot be created between heading and destination content. It is unlikely to be of much help to the user wishing to look up a restaurant guide in English

As discussed in the chapter on cohesion in navigation, a link, in this case a teaser heading, cannot function cohesively where the user is led to ‘pre-suppose’ a certain content based on the lexis in the label text but where that label does not correspond to the actual destination content. The effect is a loss of orientation and context. This is particularly the case if the destination content is in the source language, as is the case with the example for the information on sports in Leipzig, as mentioned previously. This is not only a waste of time for the user but it also questions the credibility of a web site. At the very least, a note should be added to the link label to state that the content is available only in German. The second issue with the Leipzig example is that the teaser heading promises ‘sports highlights’ but the second link provided delivers a list of all the concerts, sports meetings and theatre events that are taking place during the year. It is a comprehensive list and is in English, as promised, but it would enhance cohesion if the sporting events were filtered out and made accessible directly from the teaser text in question.

6.1.1.5 Direct address and superfluous text in the teaser text

The use of direct address in the body text of the teaser text is a common feature in the English teaser texts on the Munich, Frankfurt, Berlin and Leipzig web sites. In particular, the imperative and the interrogative are used.

The imperative is used in various positions in the teaser texts, at the start, in the middle and at the end. The imperative at the start of the teaser body text in

Figure 6.9., ‘Learn about’, allows for the immediate introduction of the subject matter:



Figure 6.9: Imperative used in teaser text, Munich site
(<http://www.muenchen.de/Stadtleben/6941/index.html>)

Consequently, a cohesive chain is created through repetition (education, employment) and superordinate / hyponymy (education: university, training).

The imperative used in Figure 6.10 is a lengthier piece of text in comparison with the last example above, which has an effect on the focus of the text:



Figure 6.10: Imperative used in teaser text, Munich site
(<http://www.muenchen.de/home/60093/Homepage.html>)

As a result, the focus shifts slightly from the core message of the text to the user as expressed in the word ‘yourself’. The main point of the teaser is arrived at on line four with the mention of ‘this virtual tour’.

Imperatives occurring mid-text can result in repetition within a short teaser, as seen in the text below where the individual terms of the phrase ‘hotels with rooms’ are repeated on the fifth line of the text:

Accommodation



With over 300 hotels with rooms bookable now, we are sure to have something to suit you. Choose from lovely apartments or hotel rooms at unbeatable rates. [more »](#)

Figure 6.11: Imperative in teaser text, Munich site
(<http://www.berlin.de/english/>)

In many teaser texts in the corpus web sites the interrogative occurs mainly but not exclusively at the start of the body text and sometimes accounts for over 50% of the entire body text, as seen in the following example from the Munich site:



Hotel Reservation Online
Would you enjoy room service in a four star hotel set amidst the luxurious scenery of Munich, or a charming bed and breakfast with friendly hosts in a typical bavarian atmosphere? Then Munich is your answer. Use our search engine to make a reservation for your stay in Munich. [»](#)

Figure 6.12: Interrogative in teaser text, Munich site
(http://www.muenchen.de/Tourismus/Accommodation_Hotels/7604/index.html)

The focus of this teaser is initially the user as in ‘Would you enjoy ..’. This is followed by information that is related to the core message but that can nonetheless be termed ‘promotional’. Phrases such as ‘luxurious scenery of Munich’, ‘charming

bed and breakfast’ and ‘typical Bavarian atmosphere’ are some such examples. The user is also addressed in the second and third sentences in this teaser, firstly using a declarative ‘Munich is your answer’ and then through the imperative ‘Use our search engine ..’. The information about the destination content that has been flagged in the heading ‘Hotel Reservation Online’ is mentioned again in the sixth line of the teaser, as in ‘ ... make a reservation for your stay in Munich’.

The phrase ‘Here you can find’ is a common feature in the initial position in English teaser texts on the corpus web sites:

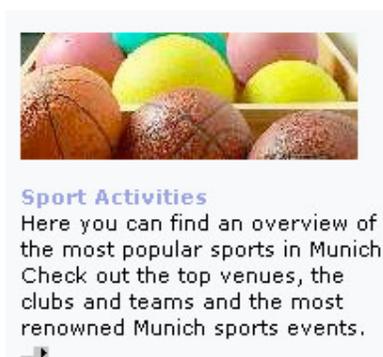


Figure 6.13: Happy Talk in a teaser text, Munich site
(http://www.muenchen.de/Tourismus/Sport_Fitness/78855/index.html)

The text above also provides the information that an ‘overview of the most popular sports in Munich’ is to follow. In addition, it is not clear whether the text refers to participant sports or the sporting events that involve teams such as Bayern München, where the general public is spectator but not participant.

The teaser in Figure 6.14 below is from the Frankfurt site and is a further example of text in a teaser text that is not absolutely necessary for a quick understanding of the

content when the user is scanning the page:



Figure 6.14: Happy Talk in a teaser text, Frankfurt site
(<http://www.frankfurt.de/sixcms/detail.php?id=317591>)

The core message that the teaser links to information about ‘room’, ‘price category’ and ‘location’ is found on the last line of this teaser text.

6.1.1.6 Issues with direct address and superfluous text in teaser texts

The purpose of the imperative in these teaser texts is similar to its use in English advertising texts in the print medium, where the advertiser supposedly communicates directly with the person reading the advertisement, exhorting the reader to undertake a specific action, for example, buy a certain product (Dyer 1982 p144). Depending on the verb chosen, the use of an initial position single-word imperative can be an effective tool in a teaser text environment as it allows for the speedy introduction of the subject matter, therefore, setting up the possibility of a strong link back to the heading and forward to the destination content. In particular, terms such as ‘Discover’ or ‘Explore’ in the context of tourist information are evocative and serve to awaken the user’s curiosity and encourage him/her to click and access the content flagged.

Where an imperative is quite long, however, there can be a delay in delivering the core message in a teaser text which may then not be read by the user. Where an

imperative is used mid-text, it is often necessary to repeat some of the text. It could be said that cohesion is enhanced through the lexical repetition. However, the overall effect is counteracted by the fact that the text has become unnecessarily long and may, therefore, not be read in full by the user. In addition, a teaser that consists of more than one sentence or clause can give the overall impression of fragmentation and could discourage the user from reading past the first 'fragment'.

The general wisdom of using an interrogative in a teaser text must be questioned given that the general purpose of a teaser is to provide information and not to request it. The interrogative in the teaser resembles what Kruger refers to as 'Happy Talk'. Happy Talk refers to elements of text on a web site such as 'Welcome text', 'Introductory text' or 'Explanations of what's on the site' (2000 p46). Generally speaking, such text provides no information of substance to the user. In addition, it can happen that the main message, which will normally contain the keywords for the destination content, will be pushed to the end of the text where it may not be seen by the user. By posing the question 'Would you enjoy room service in a four star hotel ...', the teaser text also adopts promotional language. A teaser, however, is not the ideal vehicle for such text due to the time element involved and the need for the text to be as precise and concise as possible. The user simply wants to have the facts about the destination content, so that a decision can be made quickly as to whether to click the teaser and access this content or not. Once again, if the findings of Eyetrack III (Outing and Rual 2004) are applied, which suggest that the user often only reads the left third of a teaser text and sometimes only the first line of the body text of the

teaser, Figure 6.15 is a reasonable view of what the user actually sees when scanning the page:



Figure 6.15: Text visible as per Eyetrack III, Munich site
(http://www.muenchen.de/Tourismus/Accommodation_Hotels/7604/index.html)

Apart from the information carried in the heading, which is a precise and accurate indication of the destination content and contains the keywords 'hotel' and 'reservation', the relevant information contained in the body text may not be read by the user as it has been pushed to the end of the text where it serves no purpose. This interrogative changes the focus of the teaser which appears to concentrate on all things but the core message. A further serious issue with the example above is that the user may assume that the teaser only links to information on 'four star hotels' that provide 'room service' or some type of 'luxury' accommodation. The fact that the online booking form is for all types of accommodation and budgets is not evident from the above text, consequently, the user could decide that the destination content of this teaser is simply not what s/he is looking for, despite the fact that it caters for

all budgets and needs. Ultimately, the interrogative serves only to lengthen a text unnecessarily and in actual fact obscure the real message of a teaser.

The inclusion of superfluous text is a common occurrence in the translated text on many of the web sites in the corpus, the teaser headed 'Hotel reservation' in Figure 6.14 being a case in point. The body text of this teaser contains four lines and 20+ words of which about five words adequately reflect the destination content. The two passages of text 'you will definitely find a suitable room' and 'the reservation system of Tourismus+Congress GmbH' are examples of text that add nothing of substance to the information required in the teaser. In this case, the translator has remained faithful to the ST, which illustrates the danger of following a ST closely without giving due consideration to the requirements of text in an online environment, that is in terms of reading patterns and the importance to the user of precise and concise text that includes keywords. It also underlines the issues that arise where the content and the structure of a ST that has not been optimized for cohesion are followed to the letter in the translated text.

6.1.2 Widgets

The widget chosen for analysis in this study is an online accommodation booking form. It is the type of widget that is available on each of the web sites in the corpus and is also one with which people are generally very familiar today.

The purpose of the online hotel reservation widget is to enable the user firstly, to carry out a search for a specific type of accommodation based on location, price category and facilities and, secondly, to book the accommodation that is proposed.

As such, the widget is a significant cohesive device in the provision of this online booking service. The cohesive capability of a widget depends on it being in the correct language, and on the field labels (e.g. ‘room type’ or ‘no. of nights’) and the options offered within these fields (e.g. ‘double room’ or ‘twin room’) being unambiguous, precise and accurate. The results provided for a search for accommodation on the corpus web sites do not form part of this study as, at the time of data collection, the information returned in such searches for accommodation were, for the most part, from external sites over which the web writer/translator had no influence. Two findings in particular are relevant in the context of cohesion, widgets and the translated web site:

- Many of the labels in the English online booking forms on the corpus web sites contain terms that are in the SL or that are in the TL and are ambiguous, incomprehensible or inaccurate.
- Lists that are part of the widgets and that contain additional options for booking accommodation are often randomly ordered and therefore lacking in consistency and cohesion.

6.1.2.1 Widget labels in the SL, ambiguous, imprecise or random

The choice of field label, i.e. a label used as a superordinate for a list of items such as room types, and the options listed in these fields are problematic on a number of the corpus web sites. One such example is the field label used for the list of types of accommodation, which varies from web site to web site. Accommodation type is referred to as ‘provider type’ in the Munich widget, ‘Typ’ in the Münster one, while

Hamburg refers to ‘Hoteltype’ and Garmisch-Partenkirchen to ‘Type of establishment’.

In terms of field label, the Münster widget in Figure 6.16 below is of particular interest:

The screenshot shows a search widget titled "Accommodations in Münster". Below the title is a paragraph: "This booking platform triggers a list of available accommodations in Münster and enables direct online booking. If you need assistance, do not hesitate to contact our reservation center by phone or mail." Below this are several search criteria, each with a dropdown menu: "Situation" (set to "All"), "Typ" (set to "All"), "Number" (set to "1") and "Category" (set to "All"), "Arrival" (set to "16", "4", "2007"), and "Nights:" (set to "1"). At the bottom of the widget is a red button with white text that says "Search available accommodations" followed by a right-pointing arrow. Below the widget are two blue links: "Extended Search" and "List of all accommodations".

Figure 6.16: Hotel reservation widget, Münster site
(<http://www.muenster.de/stadt/tourismus/en/hotels.html>)

In the widget in Figure 6.16, the labels ‘Arrival’ and ‘Nights’ refer to the date of arrival and the number of nights the party wishes to stay in the chosen accommodation. The term ‘Situation’ refers to the ‘location’ of the hotel, for example ‘Münster – Center’ or ‘Münster – Angelmodde’. ‘Typ’ refers to the ‘type of accommodation’. There are two possible meanings for the field labelled ‘number’ in this context. It could refer to the number of persons in the party or it could be the number of rooms required. ‘Category’, in this case, refers to the type of room as in ‘double room’ or ‘single room’.

If the ‘Extended Search’ button is clicked, a different widget is opened that is

visually very different to the original one:

Service: **Hotels (reservations)**

Accommodations in Münster

Arrival: 4, 3, 2007 Rooms: 4, All

Duration (days): 1 night Budget: Euro

Category: All

Typ: all Name of accommodation: _____

Location: All Street name: _____

Restaurant Conference Rooms Family run accommodation

Elevator Bar Minibar

Parking Pets Internet Connection in Room

Garage Biks parking Wireless LAN

Nonsmoking Rooms Pool Sauna

cable/Satellite TV

Sort: Rate Results on page: 10

Figure 6.17: Extended hotel reservation widget, Münster site
(<http://www.muenster.de/stadt/tourismus/en/hotels.html>)

The widget in Figure 6.17 also contains input-boxes in addition to drop-down menus and has a list of optional extras indicated both textually and in iconic form that can be ticked by the user. The field terms used in the extended widget are not identical to those in the original widget. ‘Category’ now refers to the star rating of a hotel and not the room type while ‘Duration (days)’ refers to the number of nights to be spent in the accommodation. ‘Situation’ has been replaced by ‘Location’.

In general on the corpus web sites, the labels used to describe the items listed, for example, on the ‘type of accommodation’ options vary widely from one web site to another and are frequently in the SL. Münster lists the types of accommodation

available as ‘Hotel, Hotel Garni, Apartments’ while Munich has ‘Hotel, Hotel Garni, Pension’. Meanwhile, the user on the Garmisch site first chooses between ‘Rooms’ or ‘Vacation Apartments’. If the user chooses ‘Rooms’, the types of accommodation available are ‘Hotel, Gasthof, Gästehäuser, Privat’. However, if s/he chooses to carry out a search for ‘Vacation Apartments’, the choice is ‘Ferienhäuser’ or ‘Ferienwohnungen’.

The widgets listing the types of rooms available show a similar problem. For example, the options offered in the ‘room type’ list on the Munich site as shown in Figure 6.18 below are a mixture of English, German and French, and include German abbreviations, for example ‘DZ zur Alleinben.’:



Figure 6.18: Hotel reservation widget, list of room types, Munich site
(http://www.muenchen.de/Tourismus/Accommodation_Hotels/hotel_reservation/76390/index.html)

The Berlin hotel booking form is accompanied by a map of Berlin that displays some of the main landmarks in the city, with the aim of giving the user a sense of

location when booking accommodation in the city:

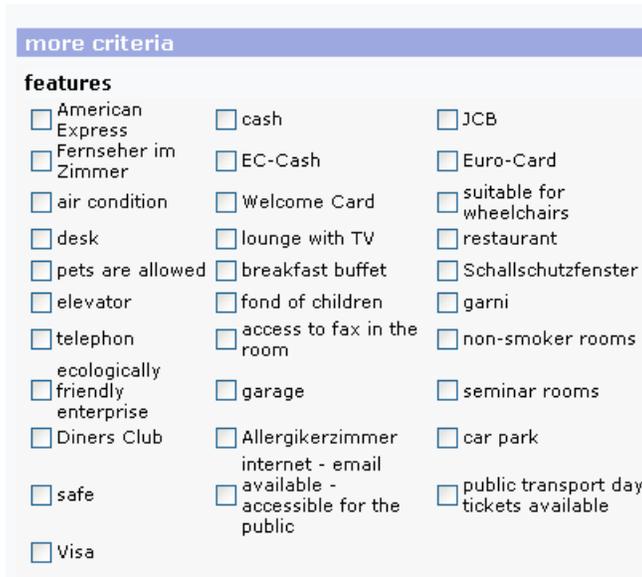


Figure 6.19: Map of Berlin with hotel reservation widget, Berlin site
(<http://www.berlin.de/english/accommodation/hotels/index.php>)

A number of items such as ‘Friedrichstrasse’ or Ku’damm’ are quite properly in the SL. However, the map also includes text such as ‘ENTF. HBF CA. 12 KM’ that is not only in abbreviation but also in the SL.

In addition to containing the acronym, JCB, the list of ‘more criteria’ on the Munich site contains a number of other items that are problematic in terms of cohesion, for example the SL terms ‘Schallschutzfenster’ and ‘Fernseher im Zimmer’ and the TL

terms ‘Euro-Card’, ‘EC-Cash’ and ‘cash’:



The image shows a screenshot of a web interface titled "more criteria" with a sub-section "features". It contains a list of 25 items, each with an unchecked checkbox. The items are arranged in three columns:

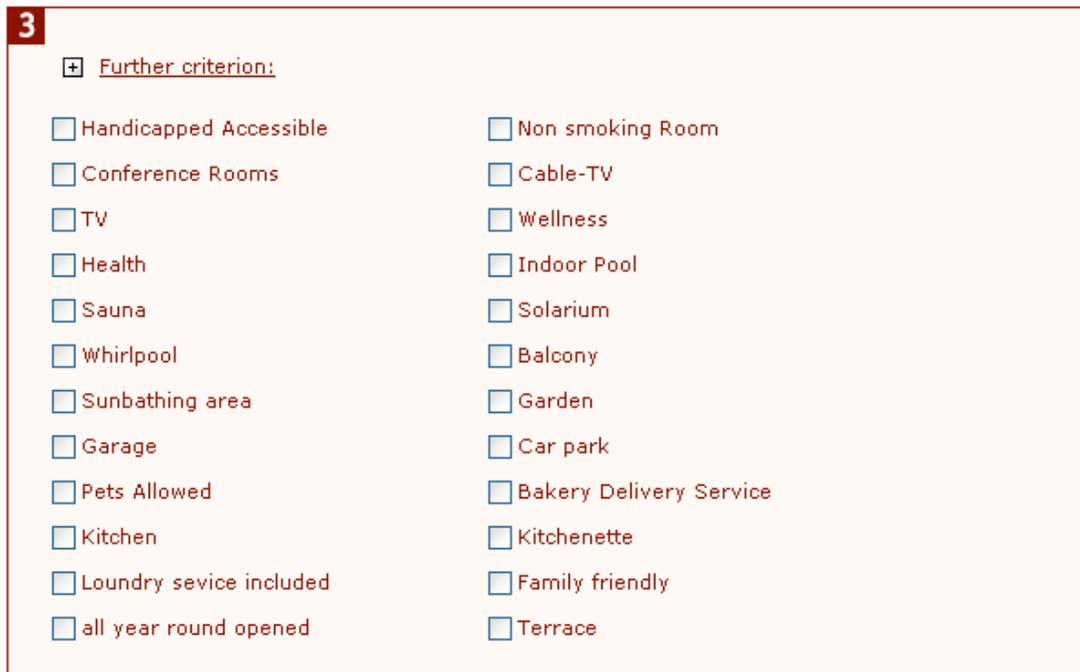
Column 1	Column 2	Column 3
<input type="checkbox"/> American Express	<input type="checkbox"/> cash	<input type="checkbox"/> JCB
<input type="checkbox"/> Fernseher im Zimmer	<input type="checkbox"/> EC-Cash	<input type="checkbox"/> Euro-Card
<input type="checkbox"/> air condition	<input type="checkbox"/> Welcome Card	<input type="checkbox"/> suitable for wheelchairs
<input type="checkbox"/> desk	<input type="checkbox"/> lounge with TV	<input type="checkbox"/> restaurant
<input type="checkbox"/> pets are allowed	<input type="checkbox"/> breakfast buffet	<input type="checkbox"/> Schallschutzfenster
<input type="checkbox"/> elevator	<input type="checkbox"/> fond of children	<input type="checkbox"/> garni
<input type="checkbox"/> telephon	<input type="checkbox"/> access to fax in the room	<input type="checkbox"/> non-smoker rooms
<input type="checkbox"/> ecologically friendly enterprise	<input type="checkbox"/> garage	<input type="checkbox"/> seminar rooms
<input type="checkbox"/> Diners Club	<input type="checkbox"/> Allergikerzimmer	<input type="checkbox"/> car park
<input type="checkbox"/> safe	<input type="checkbox"/> internet - email available - accessible for the public	<input type="checkbox"/> public transport day tickets available
<input type="checkbox"/> Visa		

Figure 6.20: Hotel reservation widget, additional resources, Munich site
(http://www.muenchen.de/Tourismus/Accommodation_Hotels/hotel_reservation/76390/index.html)

In addition, the items on the list are randomly placed and not ordered according to subject area, for example.

Consistency is fundamental to cohesion in that it enables the user to presuppose the next item in a series, for example as part of a list of items, and to carry out the intended task as quickly as possible. In terms of consistency, the accommodation widget on the Garmisch-Partenkirchen site takes a somewhat unique approach. When booking accommodation on this site, the user must first choose between ‘Rooms’ and ‘Vacation Apartments’ (sic). Once this first choice is made, a list of

options referred to as ‘Further criterion’ is available:



3

Further criterion:

<input type="checkbox"/> Handicapped Accessible	<input type="checkbox"/> Non smoking Room
<input type="checkbox"/> Conference Rooms	<input type="checkbox"/> Cable-TV
<input type="checkbox"/> TV	<input type="checkbox"/> Wellness
<input type="checkbox"/> Health	<input type="checkbox"/> Indoor Pool
<input type="checkbox"/> Sauna	<input type="checkbox"/> Solarium
<input type="checkbox"/> Whirlpool	<input type="checkbox"/> Balcony
<input type="checkbox"/> Sunbathing area	<input type="checkbox"/> Garden
<input type="checkbox"/> Garage	<input type="checkbox"/> Car park
<input type="checkbox"/> Pets Allowed	<input type="checkbox"/> Bakery Delivery Service
<input type="checkbox"/> Kitchen	<input type="checkbox"/> Kitchenette
<input type="checkbox"/> Laundry service included	<input type="checkbox"/> Family friendly
<input type="checkbox"/> all year round opened	<input type="checkbox"/> Terrace

Figure 6.21: Hotel reservation widget, additional resources, Garmisch-Partenkirchen site

(<http://www.garmisch-partenkirchen.de/en/dbdbfbf19-d2d0-4d53-027a-1694e5071f61.html?ggsuchaction=suche&action=&channel=html&lang=en&imxsession=c14a39a4b89ef3bd3b6009e99a87104a&artid=%7Bdbdbfbf19-d2d0-4d53-027a-1694e5071f61%7D&quartier=zimmer&volltextsuche=Name#>)

Some of these options are available for both ‘Rooms’ and ‘Vacation Apartments’ (sic), while others are available for either one or the other. Items such as

‘Kitchenette’ and ‘bakery delivery service’ apply only to the self-catering option.

‘Sauna’ and ‘Indoor Pool’ are available only for the hotel option. A search for accommodation will return no results, for example, where the user books a self-catering apartment and ticks the option ‘Sauna’. The same applies if a hotel room is booked and ‘bakery delivery service’ is ticked.

6.1.2.2 Issues with widgets in the SL, ambiguous, imprecise and random

A basic requirement of a widget in order to function as a cohesive device is that it can be recognised by the user as being in the appropriate language, that is English in this case. SL items among a list of TL items will not contribute to cohesion but will prevent the user from being able to carry out the task that s/he set out to achieve, which in this case is to book accommodation.

As with most things in an online environment, precision and speed play an important role in enabling the user to carry out specific tasks. In the case of widgets, this means that field labels such as the ones in the accommodation widgets must be accurate and unambiguous if they are to serve the purpose for which they have been created. Field labels should state clearly the type of information that is listed for that label and the options in the list must be equally unambiguous. Such field labels must be clearly a superordinate of the list of items that they represent. Terms such as 'situation' or 'provider type' provide no indication as to what they stand for and provide no contextual clues. They leave the user no option but to either click in order to establish exactly what these labels refer to or to leave the site altogether.

This problem is compounded on the Münster site by the fact that the word 'category' has one meaning on the initial hotel booking widget and an entirely different meaning when used in the extended search features list. In order to function cohesively such labels must be not only specific but also consistent. In tourist accommodation literature, there are standard terms that function as superordinates for very specific hyponyms. One such example is 'star-rating' or 'rating', which refers to the classification of an accommodation expressed in stars. If the label does

not immediately suggest to the user what s/he can expect on clicking such a field label, the text in such a drop-down menu or input field cannot function cohesively. Terms that are ambiguous or imprecise can have the same effect as those that are in the source language. If they are not an accurate description of what they represent they will not be recognised by the user, they cannot function cohesively. Consequently, the user cannot access the information or service offered on the web site.

The use of ambiguous language may not only lead to a lack of cohesion in a widget but could also cause offence. Figure 6.22 shows one of the items on the extended criteria lists on the Garmisch site reads:

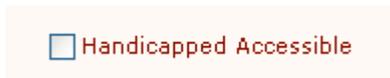


Figure 6.22: Item in extended list, hotel reservation widget, Munich site
(<http://www.garmisch-partenkirchen.de>)

while Munich lists an item with the text:

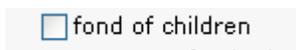


Figure 6.23: Item in extended list, hotel reservation widget, Munich site
(<http://www.muenchen.de/>)

Acronyms or abbreviations in one's native language are often difficult enough to decipher but will generally not be understood at all in a language that is not understood by the user and will not be capable of functioning cohesively through repetition or superordinate/hyponym. The information offered on the Berlin site in terms of location and distance from the main points of arrival in the city, e.g. airport, main railway station, could be very useful to the user booking accommodation.

However, in the form that this information is in, it is unlikely to be able to contribute to cohesion in this case.

Acronyms have different meanings in different countries. For example, a 'JCB' in an English-speaking country could refer to a mechanical excavator or it could also be the 'Japan Credit Bureau' which issues the JCB credit card. While it is unlikely to be the former on a hotel booking widget, it is not clear whether it is actually the 'Japan Credit Bureau'

The list of 'further criteria' in a hotel booking form is very useful for allowing the user to narrow a search for accommodation by choosing exactly which facilities s/he wishes to have in the chosen accommodation. The lack of cohesion on such lists caused by the lack of consistency and logical ordering of the items forces the user to read each individual item carefully instead of simply scanning the list to find groups of related items that are relevant to the user requirements. A more logically ordered list, where items that belong in the same semantic category are grouped together would enhance the cohesive effect of such lists.

In terms of a widget for booking accommodation, cohesion is enhanced where options or facilities that are listed can actually be offered as part of the booking that is made by the user. In the case of the Garmisch-Partenkirchen widget, it is possible that due to the structure of the widget and the choices that are possible for one type of accommodation and not for another, a lot of custom is lost as the user is unaware that certain services or options are not available for all types of accommodation.

6.2 Conclusion

Chapter VI presented the findings of the analysis of cohesion in Page Content, and specifically teaser text and widgets. It was shown that the creation of a cohesive link through repetition, linking the teaser heading, the body text and the destination content is possible where these elements of text contain a keyword or keywords that give a concise and precise indication of the destination content and are in the appropriate language. It was also seen that superfluous text in a teaser text can prevent the user from reading important keywords in a body text of the teaser and creating a link back to the teaser heading and forward to the destination content. With widget lists, cohesion created through superordinate and hyponym is only possible where the labels used for superordinate and hyponym are precise, of the correct level of specificity and in the appropriate language.

As with the issues listed in Chapters IV and V, recommendations on some of the issues outlined in Chapter VI are provided in the final chapter of this dissertation. A checklist for addressing specific issues in cohesion is included in Appendix B. It can be used either as a pre-translation tool or for a web site that has already been translated.

This final chapter also provides a summary of the aims and achievements of the study. It discusses the role of the freelance translator in web site translation and makes suggestions for the incorporation of web site translation in translator training. Suggestions for further research are made.

Conclusion

Conclusion

7.0 Introduction

This concluding chapter reviews the aims and achievements of the project. It discusses the role of the freelance translator in web site translation and makes suggestions for the training of web site translators. Some examples of the information available to the web site translator are outlined in Appendix A. This chapter outlines recommendations for a number of the issues that emerged in the analysis of the corpus web sites and supplements this with a check list provided in Appendix B. Suggestions for further research are made.

7.1 Aims and achievements of study

This study, which was inspired by the researcher's experience as a freelance translator and as Chairperson of the Irish Translators' and Interpreters' Association, set out to investigate the issue of cohesion in the TL content of multilingual web sites. It was hypothesized that the hypertextual nature of web sites and the Web in general has led to the need to examine and possibly re-define the concept of cohesion in an online environment. It was further hypothesized that the implications of a lack of cohesion in an online environment are more critical than would be the case with the printed medium.

Traditional models of cohesion - Halliday and Hasan (1976), Hoey's model of lexical cohesion (1991), Martin and Rose (2002) and Halliday and Matthiessen (2004) were examined. It was established that continuous text on a web site could be analysed using such models. However, it must be borne in mind that such text is part of a network comprising a web page, web site and Web, and as such may, for example, have been optimized for online search, in which case the cohesive devices used within it and the frequency with which they are used may differ from those seen in a comparable text in the print medium.

As regards cohesion involving other instances of text on a web site, it was found that not all aspects of traditional approaches to cohesion can be applied in their entirety, given the fact of hypertext, the manner in which the user finds and accesses Web content and the resulting reading paths in an online environment. The traditional lexical cohesive devices of 'exact repetition' and 'simple repetition' were found to be important cohesive devices in an online environment, especially where they are expressed as keywords or keyword phrases. Web site-specific collocation was also seen to contribute to cohesion.

Three key areas of user interaction with a web site were identified: *Online Search*, *Navigation* and *Page Content*. They reflect how the user finds and accesses Web content using a search engine, how the user navigates a web site, moving from page to page, and between web sites, and finally how the user interacts with on-page elements of web site text.

As lexical cohesion appeared insufficient on its own to account for cohesion on a web site, aspects of visual composition were examined for their relevance to

cohesion on a web site. It was established that, for example, the cohesive device of repetition in an online environment is reliant on aspects of visual communication in order to be capable of functioning cohesively. Significantly, not only was the concept of ‘framing’, which is comparable with cohesion in the textual composition, deemed relevant to web site cohesion. The two additional elements of the compositional metafunction, ‘information value’ and ‘salience’, also play an important role in web site cohesion.

A definition of cohesion in a web site was proposed as a starting point for the analysis of web site cohesion:

Web-specific cohesion refers to the network of lexical items, referred to as specifically-defined keywords that link various parts of a web site in a multi-linear fashion. These lexical items are not cohesive *per se* but rely on aspects of visual composition in order to function cohesively.

Crucially, lexis that is capable of functioning cohesively on a web site reflects the language of the Web user.

As no model existed specifically for analyzing cohesion on the translated web site, a model was created which builds on the concept of lexical cohesion and draws on aspects of visual cohesion. The model is also informed by aspects of Search Engine Optimization, Information Architecture and Web Usability. The model was tested on a corpus consisting of the official web sites for ten cities in Germany, significant findings presented and the implications for the translator discussed.

This study has made a contribution to Translation Studies on a number of levels. On

a theoretical level it has added to the body of research already available in particular in three ways:

- It has developed a definition of web site cohesion in the context of the translated web site which draws on aspects of traditional models of cohesion in addition to elements of visual communication.
- It has identified key issues in relation to the translation of web sites that are of importance not only to freelance translators but also to those involved in larger web site localization projects.
- It has developed a model for the analysis of aspects of cohesion on the translated web site which can be applied to web sites other than those that make up the corpus.

On a practical level, this study has also made a contribution in a number of ways:

- It has provided an insight into the nature of the work carried out by the freelance translator in web site translations and has described the changing role of the translator in this area.
- It makes suggestions for the training of translators in the translation of web sites
- It makes recommendations, supported by the check list in Appendix B, on how to address many of the issues that are outlined in the analysis of the corpus web sites as regards Online Search, Navigation and Page Content.

7.2 The role of the web site translator / translator training

The following suggestions on the role of the translator in web site translation, which incorporate suggestions for translator training, relate specifically to the role of the

translator in the creation of a cohesive web site in translation. It does not provide a detailed account of how the translation will be carried out or with which programs, CAT tools or file formats. It addresses the translation as ‘product’ and concentrates on the challenges facing the web site translator and the knowledge and skills required to create a cohesive web site. The issues discussed here will inform new approaches to the training of translators in web site translation.

It should also be noted that this discussion of the role of the translator represents the ‘ideal’ situation, which will inevitably differ from one situation to another, depending on factors such as the overall client attitude and approach to the translation of his/her web site and translator experience.

Moving away from the concept of a text with one beginning and one end, the translator requires a good understanding of the networked nature of the WWW and its web sites and of the workings and consequences of hypertext, i.e., that each page is linked to all other pages on a web site and that reading paths are multilinear.

A short, non-technical and informative online tutorial on how web sites are designed and how the different elements of a web site work together would be sufficient, in addition to an increased awareness of how the translator interacts with web content him/herself.

It is important to be familiar with how users access web content (primarily using a SE), what the user sees when a list of search result appears on the computer screen and, particularly, how the translator can influence the text provided in a search result

(and thereby influence, to an extent, the overall ranking of a web page with the SEs).

An understanding of the significance of deep-linking is required. It is becoming increasingly important as it takes the user directly to the content required, thereby avoiding unnecessary navigation of the web site but placing demands on web site cohesion in terms of context and orientation.

There are a number of online tutorials on SEO for those learning about the concept for the first time. The recommendations are straightforward and easy to follow. It is important for the client to allow the translator to translate the tags and metatags required, even though they may not have been optimized for the source web site.

In tandem with SEO, the translator must understand the significance of keywords and keyword phrases on a web site, how vital they are to cohesion on a web site and the significant role of user language when choosing keywords for a web page. This is particularly important in translation as the concept of equivalence, mentioned previously (s. 2.2.2.1), of necessity includes user language.

It is not sufficient to simply 'translate' keywords as if they had no role to play beyond the immediate text in which they are anchored. Keywords must be researched, defined and tested until the most appropriate ones are found for the content in question. The web page must be viewed as a network and the keywords must be placed there strategically both for the web spider and the user. Keywords software such as GoodKeywords or Google Adwords can be used to research the best options for the translated text, as their statistics on keywords are based on actual searches using those keywords. The program Widexl can be used to test the value of

the keywords with reference to the general content of the page for which they have been defined. The translator can learn to access and read web site statistics which show, among other things, the keywords and keyword phrases used to access a site, which search engines are directing users to the site and with what frequency, the top ten countries accessing the site etc. These statistics can also indicate whether the keywords that are defined for the site are actually those being used by the user in online search. Appendix A shows some examples of web statistics (with the kind permission of the Executive Committee of the ITIA). The translator could learn to monitor the competition web sites, that is the top ten web sites in a SERP that are returned for the same keyword(s), and to examine how those sites are optimised for keywords and tags.

The translator should understand the significance of the navigation labels on a site, i.e. that link labels are firstly significant in terms of SEO but also that they refer to the page to which they link. It is important to achieve a cohesive link through repetition between, for example, a link label, the keywords in the destination page heading and in the destination page title. A good exercise here is to work through multilingual web sites and see at first hand how they function cohesively or if not, to examine where the problem lies.

The freelance translator, and indeed the translation company, often has little influence over the text once it has been submitted to the client, despite the considerable responsibility of creating a cohesive text in an online environment. In

this respect Minogue (2008) emphasises the importance of testing the multilingual content on the web site before it goes live. This could help to avoid situations such as that which pertains on the Regensburg site in the corpus where the user is not linked directly to the TL content indicated by the flag icon. This would also help to ensure that a TL link does have TL content as destination.

The role of the translator would not normally extend to issues of web design. However, if an important element on a page, such as a TT indicator, will possibly not be seen by the user, then it should be pointed out to the client.

On a practical level, the freelance translator must understand how to cost the translation of a web site. Word count as a basis for the translation fee is not sufficient if the translator is going to address issues such as SEO.

7.3 Addressing issues of cohesion in the corpus web sites

7.3.1 Cohesion and Online Search

Where the main keyword of a source web site has a high ranking with a search engine, a similar ranking should be aimed for in the TT. The high ranking of the main keyword should be exploited for the secondary keywords for a page and user language should be taken into consideration when identifying keywords. With the example of the Munich ‘restaurants’ section, the keyword ‘restaurant’ should be placed strategically on the page. The page title for this section should include at least the words ‘munich’ and ‘restaurant’. The keywords should be placed in order of

importance, starting with the most important on the left-hand side. The text 'muenchen.de' should be removed from all page titles in the TL content as this SL term cannot create a cohesive tie with TL terms used in a search query. The description content metatag for the main page in the restaurants section should provide a concise and precise description of the page and should contain two or three of the most important keywords (i.e. 'munich' and 'restaurant').

It is unclear which criteria Google applies when deciding whether to use the defined description content metatag for its search results. However, it would appear that the likelihood is greater that it will be used if it is the correct length, does not contain the word 'Welcome' and contains the most important keywords for the web page in question.

7.3.2 Cohesion and Navigation

Access to foreign-language content via a global navigation system should be indicated at least using the word for the language in the language, e.g. 'Deutsch', or perhaps together with a flag icon. If the global system is to function cohesively in this case, the user should be linked directly to the relevant TL content, without the need to navigate through other languages or irrelevant content.

The SL navigation tabs on the English Berlin pages should be removed or translated. There can be no cohesion between a link label in one language and its destination content in another language. In addition, where the link label is wrongly in the SL, a valuable opportunity of placing a keyword in the link will be lost. Thirdly, content that may be relevant to the user needs will not be accessed. It is important to ensure

that the correct link text is placed in a TL link so that it will connect to the correct TL text.

Where possible, a site-specific search result should be designed so as to look like a standard search engine search result. This would result in the optimized tags and metatags being used again. It would also aid cohesion in that the user is already familiar with such search features from online search.

It is preferable, where possible, to return only English-language results for queries in English. In addition, the headings on the search result page should be in English, for example 'Search results' instead of 'Suche'. The text 'KölnTourismus' should be removed from the link as it is bound to confuse the user and, even if it was translated, would not necessarily add any useful information.

7.3.3 Cohesion and Page Content

The example provided of the teaser heading 'Fressgass' can be translated, perhaps simply as 'restaurants', as the term cannot otherwise function cohesively in the English-language content. The 'Eating out' teaser heading on the Cologne site could function cohesively if it were 'Restaurant guide', or similar. The word 'gastro' has different connotations in English and should be avoided in the culinary world in English.

The presence of text in the teaser text that is superfluous serves only to obscure the actual message of the teaser and should be removed.

The main keywords should be placed at the start of the body text, as it is possible that the user will only read the first few words or the first line of the teaser text.

The text of a widget must be in the correct language or the labels and lists will be incapable of functioning cohesively. The field labels (e.g. 'room type'), must be unambiguous and clearly indicate the superordinate/hyponym relation. At the very least, the user attempting to book accommodation using such a widget could decide to abandon the web site and go elsewhere to book accommodation. In the worst case, the user could book and possibly pay for accommodation where it transpires that the type of accommodation, the location or the facilities are not what the user thought s/he was booking.

SL acronyms cannot function cohesively in the TL as they are meaningless. They should therefore be avoided. Where necessary, the text can be provided in full.

Overall, Kilian's words apply to the creation of a cohesive text on a web site: 'Every sentence, every phrase, every word has to fight for its life' (1999 pp58-9).

7.4 Suggestions for further research

The model for the analysis of cohesion in the translated web site can be replicated and applied to other types of web site in order to test its viability.

Web site cohesion is an under-researched area and would, therefore, benefit greatly from research into the nature of overall cohesion on a web site. In particular, the issue of cohesion and continuous text on a web site requires further research in terms of its role within the hypertextual system of the surrounding text. The findings in

this study will have implications that require research both for the initial training of translators and for ongoing professional development.

References

Primary sources: Corpus web sites

www.berlin.de

www.frankfurt.de

www.garmisch-partenkirchen.de

www.hamburg.de

www.koeln.de

www.leipzig.de

www.muenchen.de

www.muenster.de/stadt/tourismus/index.html (parallel site to
<http://www.muenster.de/stadt/tourismus/en/>)

www.regensburg.de

www.stralsund.de

Secondary sources:

Hardcopy:

Baker, M. 1992. *In Other Words: A Coursebook on Translation*. London, New York: Routledge

Baldry, A. and Thibault, P. 2006. *Multimodal Transcription and Text Analysis*. London: Equinox

Bateman, J.A. 2008. *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. Basingstoke: Palgrave Macmillan

Berners-Lee, T. 1999. *Weaving the Web. The past, Present and Future of the World Wide Web by its Inventor*. London: Orion Business Books

Cheng, S. 2002. Globalizing an e-Commerce Web Site *IN*: Robert C. Sprung (ed.) *Translating Into Success*. Amsterdam, Philadelphia: John Benjamins Publishing Company

Crystal, D. 2001. *Language and the Internet*. Cambridge: Cambridge University Press

Dunne, K. J. (ed.) 2006. *Perspectives on Localization*. Amsterdam: John Benjamins Publishing Company

Dyer, G. 1982. *Advertising as Communication*. London, New York: Routledge

- Fawcett, P. 2003. *Translation and Language: Linguistic theories explained*. Manchester: St. Jerome
- Halliday, M.A.K and Hasan, R. 1976. *Cohesion in English*. London, New York: Longman
- Halliday, M.A.K. and Matthiessen, C. 2004. *An Introduction to Functional Grammar*. 3rd Edition. London: Arnold
- Hatim, B. and Mason, I. 1990. *Discourse and the Translator*. London: Longman
- Hoey, M. 1991. *Patterns of Lexis in Text*. Oxford, New York, Toronto, Madrid: Oxford University Press
- Hofinger, A. and Ventola, E. 2004. Multimodality in Operation: Language and picture in a museum *IN: Ventola, E., Cassily, C. and Kaltenbacher, M. (eds.) Perspectives on Multimodality*. Amsterdam, Philadelphia: John Benjamins Publishing Company
- Iedema, R. 2003. Analysing film and television: a social semiotic account of *Hospital: an Unhealthy Business IN: van Leeuwen, T. and Jewitt, C. (eds.) Handbook of Visual Analysis*. London, Thousand Oaks, New Delhi: SAGE Publications
- Jewitt, C. and Oyama, R. 2001. Visual meaning: a social semiotic approach *IN: van Leeuwen, T. and Jewitt, C. (eds.) Handbook of Visual Analysis*. London, Thousand Oaks, New Delhi: SAGE Publications
- Kilian, C. 1999. *Writing for the Web*. North Vancouver, BC: Self-Counsel Press
- Kress, G. 2003. *Literacy in the New Media Age*. London, New York: Routledge
- Kress, G. and van Leeuwen, T. 1996. *Reading Images: The Grammar of Visual Design*. London, New York: Routledge.
- Kress, G. and van Leeuwen, T. 2001. *Multimodal Discourse*. London: Arnold.
- Krug, S. 2000. *Don't Make Me Think: A common-sense approach to Web usability*. Indianapolis, Ind.: New Riders
- Landow, GP. 2006. *Hypertext 3.0. Critical Theory and New Media in an Era of Globalization*. Baltimore: The Johns Hopkins University Press
- McCarthy, M. 1990. *Vocabulary*. Oxford: Oxford University Press.

- McEnery, T. and Xiao, R. 2008. Parallel and Comparable Corpora: What is Happening? IN Anderman, G. and Rogers, M. (eds.) *Incorporating Corpora. The Linguist and the Translator*. Cleveland, Buffalo, Toronto: Multilingual Matters Ltd.
- Martin, JR. 1992. *English Text: System and Structure*. Philadelphia, Amsterdam: John Benjamins Publishing Company
- Martin, JR. and Rose, D. 2002. *Working with Discourse: Meaning beyond the Clause*. London: Continuum
- Morley, J. 2006. Lexical Cohesion and Rhetorical Structure IN: *International Journal of Corpus Linguistics*. 11 (3), pp266-282
- Nielsen, J. 1999. *Designing Web Usability*. Indianapolis: New Riders Publishing
- Nielsen, J. and Loranger, H. 2006. *Prioritizing Web Usability*. Berkeley, CA: New Riders Publishing
- O'Hagan, M. and Ashworth, D. 2002. *Translation-Mediated Communication in a Digital World*. Clevedon, Buffalo, Toronto, Sydney: Multilingual Matters Ltd.
- O'Toole, M. 1994. *The Language of Displayed Art*. London: Leicester University Press
- Pollach, I. 2003. *Communicating Corporate Ethics on the World Wide Web: A Discourse Analysis of Selected Company Web Sites*. Frankfurt: Peter Lang
- Pym, A. 2004. *The Moving Text: Localization, Translation and Distribution*. Amsterdam, Philadelphia: John Benjamins Publishing Company
- Pym, A. 2006. Globalization and the Politics of Translation Studies. IN: *META* LI, 4, pp744-757
- Ramsey, E., Ibbotson, P., Bell, J. and Gray, B. 2002. Getting 'e' real: The strategic challenges of the information age for Irish small and medium-sized enterprises IN: *E-Business Review*. 2, pp118-123
- Rosenfeld, L. and Morville, P. 2002. *Information Architecture for the World Wide Web*. 2nd ed. Sebastopol, CA: O'Reilly Media Inc.
- Scharf, F., Bell, J., Loane, S., Fletcher, R. 2004. The export problems of internationalizing SMEs: some empirical evidence using a 'critical incident' technique", IN Jones, M.V. and Dimitratos, P. (eds.). *Emerging Paradigms in International Entrepreneurship*. Cheltenham: Edward Elgar Publishing

Smith, R. and Frawley, W. 1983. Conjunctive Cohesion in Four English Genres *IN: Text* 3 (4), pp347-374.

Snell-Hornby, M. 1988. *Translation Studies*. Amsterdam: John Benjamins

Stöckl, H. 2004. In between modes: Language and image in printed media. *IN: Ventola, E., Cassily, C. and Kaltenbacher, M. (eds.). Perspectives on Multimodality*. Amsterdam/Philadelphia: John Benjamins Publishing Company

Tóibín, C. 2006. *Mothers and Sons*. London: Picador

Toury, G. 1985. A Rationale for Descriptive Translation Studies *IN: Hermans, T. (ed.). The Manipulation of Literature*. London, Sydney: Croom Helm

Van Leeuwen, T. and Jewitt, C. (eds.). 2001. Visual Meaning: A Social Semiotic Approach *IN: Handbook of Visual Analysis*. London, Thousand Oaks, New Delhi: Sage Publications

Electronic:

Amárach Consulting. 2008. Facts and Friction: Trends in Search Engine Usage and Search Engine Marketing in Ireland. [Online]. Available from: <http://www.amarach.com/assets/files/Amarach_SEM_Conference_Presentation_March_2007.pdf> [Accessed 2 July 2008].

Amberton University. 2007. Glossary of Terms: Hypertext. [Online]. Available from: <http://www.amberton.edu/VL_terms.htm> [Accessed 10 January 2007].

Ariadne. 2007. Problem with Keyword Spamming. [Online]. Available from: <<http://www.ariadne.ac.uk/issue10/search-engines/>> [Accessed 10 February 2007].

Bebo Social Environment. 2007. [Online]. Available from: <www.bebo.com> [Accessed 15 May 2007].

Berners-Lee, T. 2002. Commemorative lecture: The World Wide Web – Past, Present and Future. [Online]. Available from: <<http://www.w3.org/2002/04/Japan/Lecture.html>> [Accessed 13 April 2004].

Biz/Ed. 2006. Simple Business Website: Glossary of Terms. [Online]. Available from: <http://www.bized.co.uk/educators/16-19/business/marketing/lesson/sup_glossary.htm> [Accessed 10 May 2006].

Bowker, L. 2005. Exploring the Notion of 'Text' in the Context of Translation Technology. [Online]. Available at:
<<http://aix1.uottawa.ca/~lbowker/TRA8901-2005.ppt>> [Accessed 12 May 2005].

Buchholz, W. J. 2007. Labeling Systems. [Online]. Available from:
<http://cyber.bentley.edu/faculty/wb/printables/labeling_systems.pdf> [Accessed 10 March 2007].

Callan, B. 2005. Search Engine Optimization Made Easy. [Online]. Available from:
<www.ketec.net/SEO.html> [Accessed 20 January 2006].

Cripps, M. J. (2007). *Hypertext Theory and WebDev in the Composition Classroom*. [Online]. Available from: <<http://www.bgsu.edu/cconline/cripps>> [Accessed 17 June 2007].

CSO. 2006. Information Society and Telecommunications 2005. [Online]. Available from:
<<http://www.cso.ie/releasespublications/documents/industry/2005/ictireland2005.pdf>> [Accessed 10 August 2007].

Dennis, J. 2004. Using Hyperlinks as a Search Engine Optimization Tool. [Online]. Available from:
< <http://www.pandia.com/sw-2004/42-hyperlinks.html>> [Accessed 13 October 2004].

DMOZ. 2005. Open Directory Project: Describing Sites. [Online]. Available from:
<<http://research.dmoz.org/~gti96/ddp/07002/#what>> [Accessed 4 August 2006].

DMOZ. 2006. Open Directory Project. Berlin entry. [Online]. Available from:
<<http://search.dmoz.org/cgi-bin/search?search=berlin+official+guide+to+the+history>> [Accessed 10 February 2007].

DMOZ. 2006. Open Directory Project. Frankfurt entry. [Online]. Available from:
<<http://search.dmoz.org/cgi-bin/search?search=Frankfurt+official+news%2C+information>> [Accessed 10 February 2007].

DMOZ. 2006. Open Directory Project. Leipzig entry. [Online]. Available from:
<<http://search.dmoz.org/cgi-bin/search?search=leipzig+service+for+citizens>> [Accessed 10 February 2007].

DMOZ. 2006. Open Directory Project. Regensburg entry. [Online]. Available from:
<<http://search.dmoz.org/cgi-bin/search?search=regensburg+municipal+government>> [Accessed 10 February 2007].

Emarketing. 2006. Search Engine Jargon Glossary. [Online]. Available from: <<http://www.emarketing-strategies.com/search-engine-glossary.htm>> [Accessed 10 February 2007].

European Agency for Safety and Health at Work. 2008. Safety and Health for SMEs. [Online]. Available from <http://osha.europa.eu/en/sub/sme/> [Accessed 10 June 2008].

Expedite Email Marketing. 2007. Internet Marketing Terms. [Online]. Available from: <http://www.expedite-email-marketing.com/internet_marketing_glossary_internetmarketingtermsdefinition.htm> [Accessed 10 September 2007].

Google. 2006. Search Results Page. [Online]. Available from: <<http://www.google.com/help/interpret.html>> [Accessed 10 February 2007].

Google. 2007a. Directory 2007. [Online]. Available from: <http://directory.google.com/Top/Reference/Libraries/Library_and_Information_Science/Technical_Services/Cataloguing/Metadata/RDF/Applications/RSS/News_Readers/> [Accessed 15 August 2007].

Google. 2007b. Language Tools. [Online]. Available from: http://www.google.ie/language_tools?hl=en [Accessed 7 July 2008].

Google. 2007c. Technology. [Online]. Available from: <<http://www.google.com/technology/index.html>> [Accessed 10 February 2007].

Google. 2007d. Webmaster Guidelines. [Online]. Available from: <<http://www.google.com/support/webmasters/bin/answer.py?answer=35769>> [Accessed 10 February 2007].

Gramene Glossary. 2004. HTML. [Online]. Available from: <<http://www.gramene.org/documentation/glossary/>> [Accessed 10 October 2004].

Internet World Stats. 2008. Internet statistics by language. [Online]. Available from: <<http://www.internetworldstats.com/stats7.htm>> [Accessed 3 July 2008].

iProspect. 2006. Search Engine User Behaviour Study. [Online]. Available from: <http://www.iprospect.com/premiumPDFs/WhitePaper_2006_SearchEngineUserBehavior.pdf> [Accessed 1 October 2007].

- Ishida, R. 2007. *Text Size in Translation*. [Online]. Available from:
<<http://www.w3.org/International/articles/article-text-size> > [Accessed 15 September 2007].
- Jaha Design. 2006. Glossary: Hypertext [Online]. Available from:
<<http://www.jahadesign.com/glossary.htm> > [Accessed 10 February 2007].
- Linking101. 2006. Glossary: Keyword [Online]. Available from:
<<http://www.linking101.com/articles/terms.htm>> [Accessed 10 February 2007].
- LISA/AIIM. 2001. The Black Hole in the Internet: LISA/AIIM Globalization Survey. [Online]. Available from:
<<http://www.lisa.org.productsurvey.AIIMsurvey.pdf>> [Accessed 20 March 2004].
- Morkes, J. and Nielsen, J. 1997. *Concise, Scannable and Objective: How to Write for the Web*. [Online]. Available from:
<<http://www.useit.com/papers/webwriting/writing.html>> [Accessed 11 April 2007].
- MySpace Social Environment. 2007. [Online]. Available from:
<www.myspace.com > [Accessed 15 May 2007].
- Netcraft. 2006. Web Server Survey. [Online]. Available from:
<http://news.netcraft.com/archives/2006/11/01/november_2006_web_server_survey.html> [Accessed 5 December 2007].
- Nielsen, J. 1996. International Web Usability. [Online]. Available from:
<<http://www.useit.com/alertbox/9608.html>> [Accessed 17 January 2007].
- Nielsen, J. 1997. How Users read on the Web. [Online]. Available from:
<<http://www.useit.com/alertbox/9710a.html> > [Accessed 10 January 2007].
- Nielsen, J. 2000. Top Ten Design Mistakes of 1999. [Online]. Available from:
< <http://www.useit.com/alertbox/990530.html> > [Accessed 10 February 2007].
- Nielsen, J. 2001. Search: Visible and Simple. [Online]. Available from:
<<http://www.useit.com/alertbox/20010513.html>> [Accessed 17 January 2007].
- Nielsen, J. 2002a. Deep linking is good. [Online]. Available from:
<<http://www.useit.com/alertbox/20020303.html> > [Accessed 10 February 2007].
- Nielsen, J. 2002b. Top Ten Guidelines for Homepage Usability. [Online]. Available from:
< <http://www.useit.com/alertbox/20020512.html> > [Accessed on 10 February 2007].

- Nielsen, J. 2005. Top Ten Design Mistakes of 2005. [Online]. Available from: <<http://www.useit.com/alertbox/designmistakes.html>> [Accessed 10 February 2007].
- Nielsen, J. 2006a. F-shaped pattern for reading Web content. [Online]. Available from: <http://www.useit.com/alertbox/reading_pattern.html> [Accessed 10 February 2007].
- Nielsen, J. 2006b. Use Old Words when Writing for Findability. [Online]. Available from: <<http://www.useit.com/alertbox/search-keywords.html>> [Accessed 17 January 2007].
- Nielsen//Netratings. 2006a. Online Search February. [Online]. Available from: <http://www.nielsen-netratings.com/pr/pr_060209.pdf> [Accessed 18 February 2006].
- Nielsen//NetRatings: 2006b. Online Search March. [Online]. Available from: <http://nielsen-netratings.com/pr/pr_060302.pdf> [Accessed 10 January 2007].
- Nielsen//NetRatings. 2006c. UK Google Share of Market. [Online]. Available from: <http://nielsen-netratings.com/pr/PR_101206_UK.pdf> [Accessed 10 January 2007].
- Nielsen//NetRatings. 2007. US Search Share Rankings. [Online]. Available from: <http://www.nielsen-netratings.com/pr/pr_071204.pdf> [Accessed 4 December 2007].
- NUS. 2007. Hypertext. [Online]. Available from: <<http://www.scholars.nus.edu.sg/epace/ht/htlecture/index.html>> [Accessed 10 February 2007].
- O'Hagan, M. 2003. Training for Localization. [Online]. Available from: <http://isg.urv.es/seminars/2003_localization_online/ohagan.html> [Accessed 10 December 2006].
- Outing, S. and Rual, L. 2004. The Best of Eyetrack III. What we saw when we looked through their Eyes. [Online]. Available from: <<http://www.poynterextra.org/eyetrack2004/main.htm>> [Accessed 10 February 2007].
- Pasternack, D. 2005. Search is Dead. [Online]. Available from: <<http://www.seotoday.com/browse.php/category/articles/id/504/index.php>> [Accessed 25 October 2005].

Pierini, P. 2007. Quality in Web Translation: An Investigation into UK and Italian Tourism Web Sites IN: *The Journal of Specialised Translation*, 8 July 2007, pp85-103. [Online]. Available at: <http://www.jostrans.org/issue08/art_pierini.pdf > [Accessed 10 October 2007].

Poynter Institute. 2007. Eyetrack07. [Online] Available from: <<http://eyetrack.poynter.org/> > [Accessed 9 November 2007].

Ravichandran, V., Azhar, J. and Afendi, H. 2004. Characteristics of Written Text in E-Commerce Websites. IN: *Internet Journal of e-Language Learning & Teaching* 1(2), July 2004. 15-32.[Online]. Available from: <<http://www.eltrec.ukm.my/ijelt/pdf/Ravichandran.pdf>> [Accessed 15 May 2007].

Sandrini, P. 2005. Website Localization and Translation pp1-8. [Online]. Available from: <http://www.euroconferences.info/proceedings/2005_Proceedings/2005_Sandrini_Peter.pdf > [Accessed 10 May 2007].

Searchengines.com. 2006. HTML Titles: Page Title. [Online]. Available from: <<http://www.searchengines.com/HTMLtitle.html>> [Accessed 10 February 2007].

SearchEngineWatch. 2006. US Google Share of Market. [Online]. Available from: <<http://searchenginewatch.com/showPage.html?page=2156451>> [Accessed 9 January 2007].

SearchEngineWatch 2007. How to use HTML Metatags. [Online]. Available from: <<http://searchenginewatch.com/showPage.html?page=2167931> > [Accessed 10 September 2007].

SEO Administrator. 2007. SEO Tutorial. [Online]. Available from: <<http://seo-tutorial.seoadministrator.com/#214> > [Accessed September 17 2007].

Sheikh et al. (Interesource Group). 2003. Review of User Needs and Resources for Inter Cultural Training. [Online]. Available from: <<http://www.intercommproject.com/en/Nat%20Review%20for%20web%20-%20Ire.doc> > [Accessed 10 May 2004].

Sky News. 2007. Sky News Online. [Online]. Available from: <www.sky.com/news> [Accessed 15 March 2007].

SLS Consulting. 2007. Internet Marketing. Glossary. [Online]. Available from: <<http://www.legalinternetmarketing.com/seo-glossary.html> > [Accessed 10 June 2007].

Softnik Strategic Software. 2005. GoodKeywords Freeware Application. [Online]. Available from:

<<http://www.softnik.com/products/>> [Accessed 10 October 2005].

TILP: Localization course. [Online]. Available from:

<<http://www.tilponline.org/CLP/coursedescriptions/monterey/Translation%20and%20Web%20Localization.pdf> > [Accessed 10 October 2006].

Timberlake, S. 2000. The Basics of Navigation. [Online]. Available from:

<<http://www.efuse.com/Design/navigation.html>> [Accessed 10 February 2007].

UCC CS Department. 2007. Software Development Glossary. [Online]. Available from:

<<http://www.cs.ucc.ie/~dgb/courses/swd/glossary.html#graphicalUserInterface> > [Accessed 10 October 2007].

University of Chula. 2007. Skimming and Scanning. [Online]. Available from:

<http://pioneer.chula.ac.th/~pkanchan/html/skim.htm>. [Accessed 10 February 2007].

USD. 2007. Glossary: Search Engine / HTML. [Online]. Available from:

<<http://www.usd.edu/library/instruction/glossary.shtml>> [Accessed 10 February 2007].

USDHHS (US Department of Health and Human Services). 2006a. Research-based web design and usability guidelines. Glossary. [Online]. Available from:

<<http://www.usability.gov/pdfs/guidelines.html#1>> [Accessed 2 May 2006].

USDHHS (US Department of Health and Human Services). 2006b. Research-based web design and usability guidelines. Navigation. pp1-13 [Online]. Available from:

<<http://www.usability.gov/pdfs/chapter7.pdf> > [Accessed 2 May 2006].

USDHHS (US Department of Health and Human Services). 2006c. Research-based web design and usability guidelines. Widgets pp1-22. [Online]. Available from:

<<http://www.usability.gov/pdfs/chapter13.pdf> > [Accessed 2 May 2006].

Vaughns. 2007. Google Ranking Factors. [Online]. Available from:

<<http://www.vaughns-1-pagers.com/internet/google-ranking-factors.htm>>

[Accessed 27 January 2007].

Webopedia. 2007a. Glossary: HTML tag. [Online]. Available from:

<http://www.webopedia.com/TERM/t/tag.html> [Accessed 10 February 2007].

Webopedia. 2007b. Glossary: Menu Bar. [Online]. Available from:

<http://www.webopedia.com/TERM/m/menu_bar.html>. [Accessed 10 February 2007].

Webopedia. 2007c. Glossary: Organic Search. [Online]. Available from:
<http://www.webopedia.com/TERM/n/natural_search.htm> [Accessed 10 February 2007].

Websearchworkshop. 2007. SEO. [Online]. Available from:
<www.websearchworkshop.co.uk/> [Accessed 10 February 2007].

WebWorkshop. 2007. DMOZ. [Online]. Available from:
<<http://www.webworkshop.net/dmoz.html> > [Accessed 25 September 2007].

Whitespace Internet Strategy. 2006. Jargon Explained. [Online]. Available from:
<www.whitespace-strategy.co.uk/knowledge_phrasebook.php> [Accessed 1 August 2006].

Widexl 2007. Internet Solutions. [Online]. Available from:
<<http://www.widexl.com>> [Accessed 25 July 2007].

Wordnet Database. 2006. Hyperlink. [Online]. Available from:
<<http://wordnet.princeton.edu/perl/webwn?s=hyperlink>> [Accessed 13 September 2007].

W3C Ubiquitous Web Domain. 2006. Mobile Web Initiative Activity. [Online]. Available from: <<http://www.w3.org/UbiWeb/>> [Accessed 10 August 2006].

Yahoo!. 2007. Directory. [Online]. Available from:
<http://search.yahoo.com/search?p=web+directory&fr=yfp-t-501&toggl=1&cop=mss&ei=UTF-8&vc=&fp_ip=IE > [Accessed 12 August 2007].

4 2 Explore. Skimming and Scanning. [Online]. Available from:
<<http://www.42explore.com/skim.htm>> [Accessed 6 August 2007].

Personal communication

October 2005 to January 2008

Mr D. Kettle
CEO
Web Design Ireland.

January to June 2008:

Ms. Alison Boardman
MD
Unique Voice Translation Service
Ireland

DCU-LS
Translation and Interpreting Company
Ireland

Ms Brigitte Eymard-Duvernay
Dipl.-Dolmetscherin
Interprète de conference
Germany

Ms. Martina Minogue
CEO
ETeams Translation Company
Ireland

Fr. Carola-Heil
CEO
Fachübersetzungen
Saarbrücken
Germany

Corpus web site administrators:

Cologne:
Hr. Alexander Esser
Redaktion koeln.de
c/o NetCologne GmbH
Am Coloneum 9
50829 Köln

Frankfurt:

Fr Melanie Hartlaub
Stadt Frankfurt am Main
DER MAGISTRAT
Amt für Informations- und Kommunikationstechnik
Frankfurt am Main
Germany

Garmisch-Partenkirchen

Fr. Rappenglück
E-Marketing
Kurverwaltung Garmisch-Partenkirchen
Garmisch-Partenkirchen
Germany

Hamburg

Hr. Torralf Köhler
Leiter Marketing & Sales
hamburg.de GmbH & Co. KG
Hamburg
Germany

Leipzig

Fr. Elisabeth Hoppe
Amt für Wirtschaftsförderung
Abt. Messe, Medien, Logistik
Leipzig
Germany

Munich:

Hr. Claudio Franchi
Leiter Produktmanagement und Webservices
Portal München Betriebs-GmbH & Co. KG
Munich
Germany

Regensburg

Fr. Elizabeth Knott
Hauptabteilung Presse- und Öffentlichkeitsarbeit
Stadt Regensburg
Regensburg
Germany

Appendix A

Web site statistics

AWStats for <http://www.translatorsassociation.ie>

Search Keyphrases (Top 10) Full list			Search Keywords (Top 10) Full list		
206 different keyphrases	Search	Percent	299 different keywords	Search	Percent
itia	74	13 %	irish	159	11.9 %
bulletin	52	9.1 %	translators	131	9.8 %
irish	43	7.5 %	itia	99	7.4 %
irish translators association	36	6.3 %	association	96	7.2 %
irish translators	16	2.8 %	bulletin	61	4.5 %
irish translators and interpreters association	11	1.9 %	translation	39	2.9 %
membership	10	1.7 %	ireland	35	2.6 %
irish association of translators	8	1.4 %	of	32	2.4 %
iti bulletin	8	1.4 %	interpreters	32	2.4 %
members	8	1.4 %	and	32	2.4 %
Other phrases	300	53 %	Other words	612	46 %

1) Search keyphrases / search keywords

Countries (Top 10) - Full list				
Countries		Pages	Hits	Bandwidth
 Ireland	ie	46955	63460	194.20 MB
 Unknown	unknown	12100	18873	30.94 MB
 Spain	es	4077	5121	16.06 MB
 Germany	de	3378	4053	11.30 MB
 Great Britain	gb	3237	4618	22.89 MB
 France	fr	2997	3594	8.23 MB
 Romania	ro	1931	1977	1.44 MB
 United States	us	1820	7462	104.61 MB
 Poland	pl	798	1097	4.54 MB
 Oman	om	787	898	9.03 MB
Others		3093	6263	50.90 MB

2) Top 10 countries visiting the site

Connect to site from				
Origin	Pages	Percent	Hits	Percent
Direct address / Bookmarks	1434	66.2 %	3158	79.7 %
Links from a NewsGroup				
Links from an Internet Search Engine - Full list	575	26.5 %	596	15 %
- Google	410 410			
- Windows Live	151 151			
- AllTheWeb	4 4			
- MSN Search	3 3			
- Yahoo!	3 3			
- Google (cache)	2 23			
- Seznam	2 2			
Links from an external page (other web sites except search engines) - Full list	152	7 %	203	5.1 %
- http://en.wikipedia.org/wiki/Lionbridge	17 17			
- http://cia.bzzz.net/lionbridge_globalizacja_niskich_plac	14 14			
- http://lionbridge.zsp.net.pl	14 14			
- http://www.anarkismo.net/newswire.php	13 13			
- http://libcom.org/news/lionbridge-globalizing-low-wages-20012008	9 9			
- http://www.surveymonkey.com/s_thankyou.aspx	7 7			
- http://www.priamaakcia.sk	5 5			
- http://www.fit-ift.org/en/europe.php	4 4			
- http://zsp.net.pl/lionbridge/index.htm	3 3			
- http://www.sft.fr/page.php	3 3			
- Others	63 114			

3) Links to site from search engines and other web sites

Statistics for:
translatorsassociation.ie

Summary

When:

Monthly history

Days of month

Days of week

Hours

Who:

Countries

Full list

Cities

Hosts

Full list

Last visit

Unresolved IP Address

Robots/Spiders visitors

Full list

Last visit

Navigation:

Visits duration

File type

Viewed

Full list

Entry

Exit

Operating Systems

Versions

Unknown

Browsers

Versions

Unknown

Referrers:

Origin

Referring search engines

Referring sites

Search

Search Keyphrases

Search Keywords

Others:

Miscellaneous

HTTP Status codes

Pages not found

4) List of different statistics that can be obtained for a web site through AWStats.

Appendix B

Checklist: Addressing cohesion in the translated web site

1. Online search

Keywords:

- Have keywords been defined for all or at least the most important pages of a site, i.e. the homepage and the main section start pages?
- Are the keywords a reflection of the true content of the page?
- Are the keywords precise and unambiguous?
- Are the keywords specific rather than general?
- Are they consistent? Has the same referent been given the same text label across the entire web site and on all web pages?
- Do the keywords reflect user language?
- Have keywords been included in the most important areas on the web site, i.e. the page title, h1 header tag and links?
- Are the keywords present in the page content (in the first paragraph or two of body text)?
- Are the keywords highlighted (italics/bold)?
- Are the keywords in the correct language?
- Can the keywords be used in hyperlinks?
- Are synonyms (car/automobile), plurals (hotel/hotels) or common misspellings (Accommodation/acommodation/accomodation) included in the keywords metatag?
- How successful was the search using the main keyword for the site?
- How successful was the search using the main keyword together with secondary keywords?
- Have the keywords been placed in keywords metatags for the given page?

Page title:

- Has a page title been defined for all pages of the web site?
- Is the page title unique?
- Is the page title a true reflection of the page content?
- Is the page title in the correct language?
- Does the page title contain the most important keyword(s)?
- Does the page title contain an appropriate number of keywords?
- Does the page title begin with the most important keyword?
- Does the page title contain redundant text?
- Do the keywords reflect user language?
- Are the keywords that are in the page title also present on the web page (body text, headings, link text)?
- Is the page title of an appropriate length?

h1 header text:

- Has a h1 header tag been defined for the homepage and other important pages of the web site?
- Is the h1 header tag a true reflection of the page content?
- Is the h1 header tag in the correct language?
- Does the h1 header tag contain the most important keyword(s)?
- Does the h1 header tag correspond to the page title?
- Does the h1 header tag contain redundant text?
- Do the keywords reflect user language?
- Are the keywords that are used also present in other text on the web page?

Description content metatag:

- Has a description content metatag been defined for each page or at least for the homepage and the entry pages for each main section?
- Is the text a true description and not simply a list of keywords?
- Does it adequately reflect the content of the page for which it is defined?

- Is the description text a repetition of the page title?
- Does the description content metatag contain redundant text?
- Does the text contain the most important keywords for the page?
- Is the text optimized for skimming and scanning?
- Is the text of appropriate length?
- Is the description content metatag in the correct language?

2. Navigation

Labels for embedded navigation systems:

- Are the labels in the correct language?
- Are the labels a precise indication of the destination content? Content on many web sites is frequently updated, changed, removed and/or added to, in particular in preparation for translation for a TL audience
- Is the label text unambiguous?
- Are the labels unique? No two labels should be identical unless the destination content is identical.
- If the label text is newly-coined is the meaning clear in the context?
- Is the level of specificity correct in the link list?
- Is the register appropriate? Labels should avoid mixing registers.
- Are lists of labels consistent, for example in terms of register, granularity and comprehensiveness?
- Is consistency of terminology applied across the web site (applies particularly to labels for global navigation systems)?
- Is the presence of foreign-language content on the web site clearly signalled on all pages of the site?
- Is it clear in which language or languages content is available on the site?
- Have keyword labels been applied, where appropriate?
- Do the labels reflect the web site hierarchy? Top hierarchy labels are more general, as the content further down is more specific. Labels in a local or contextual navigation system will be more specific than those that are part of a global navigation system.

- Are long text labels optimized for scanning and skimming?
- Can generic labels such as 'click here' or 'continue' be avoided? Is the destination content of these labels apparent?
- Are anchor links used and do the labels reflect the main content of the sections they refer to?
- Do groups of labels have parallel phrasing? For example, where possible do the link labels all begin with a noun or a verb ((Nielsen and Loranger 2006 p281)?
- The labels for contextual navigation systems rely largely on the surrounding text for their context. Are they meaningful in this context?
- Is the destination content clearly expressed in the contextual navigation label?

3. Page Content

Teaser texts:

- Does the language of the teaser text correspond to that of the destination content?
- Does the teaser heading contain keywords and is it salient (e.g. in bold)?
- Is there a cohesive link through repetition between the teaser heading, the body text of the teaser and the destination content?
- Does the body text of the teaser contain keywords?
- Where the teaser text links to a navigation page that contains a number of different subjects, is it clear to which specific item the teaser refers?
- Where the teaser body text is created from the first few lines of a destination text, does it contain keywords and does it provide an adequate description or overview of the full text?

Widgets:

- Are all textual components of the widget in the correct language?
- Do all data input and selection fields have a label?
- Are the labels precise and unambiguous and do they indicate the superordinate/hyponym relation with the items in their lists?
- In widget lists, are the items random or have they been grouped according to subject matter or listed alphabetically?