



Total Recall? A Case Study of Consistency in Translation Memory

Joss Moorkens

Translation Memory (TM)

Core Assumptions:

- Reduce Costs
- Increase Productivity
- Eliminate Repetitive Tasks
- Eliminate Inconsistency

Current Case Study

Single Translation Project

2004: 70,276 words US English to German

2008: 31,681 words US English to German

Initial Findings

Inconsistencies were present.

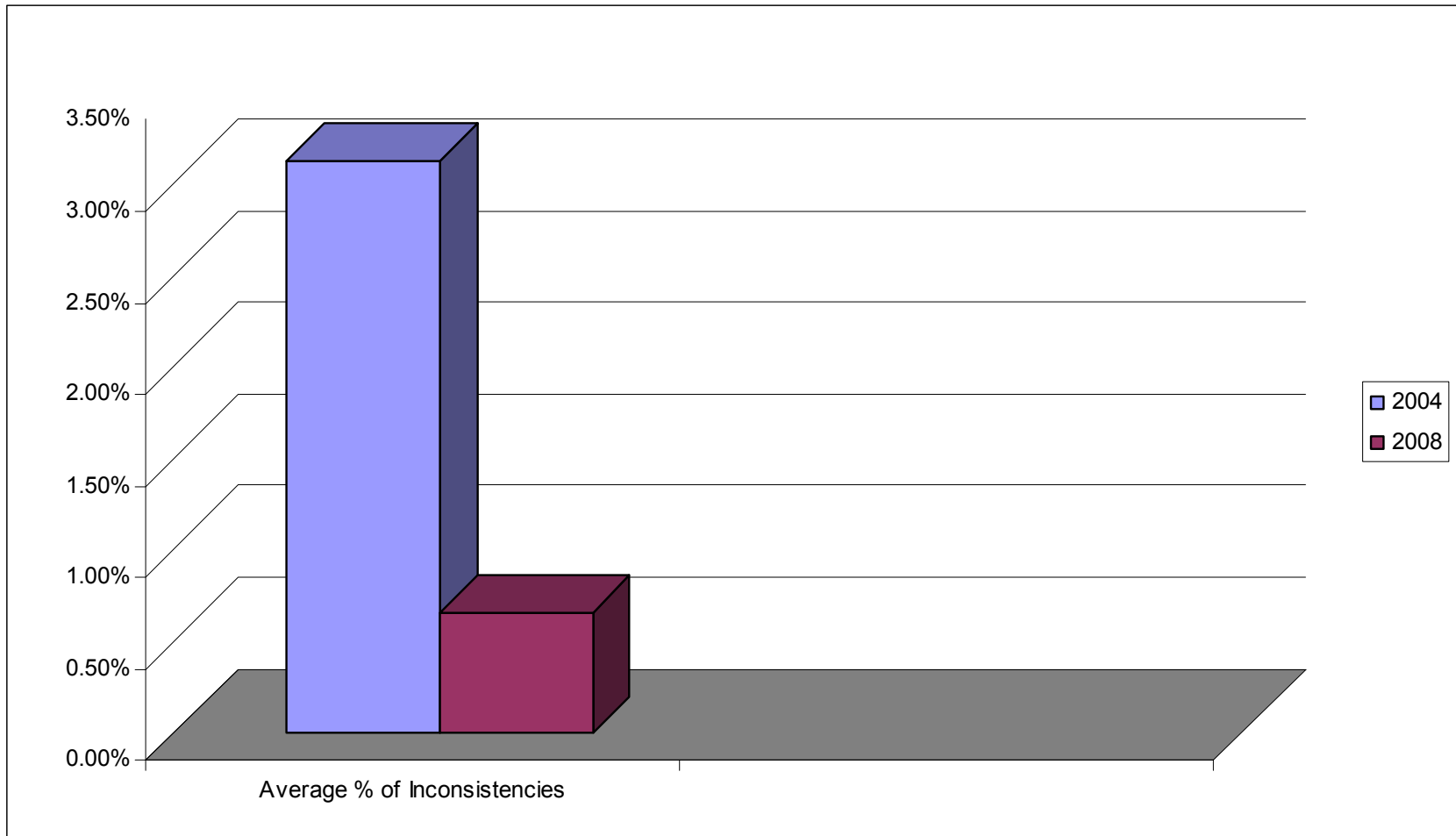
Resulting in:

Poorer Quality

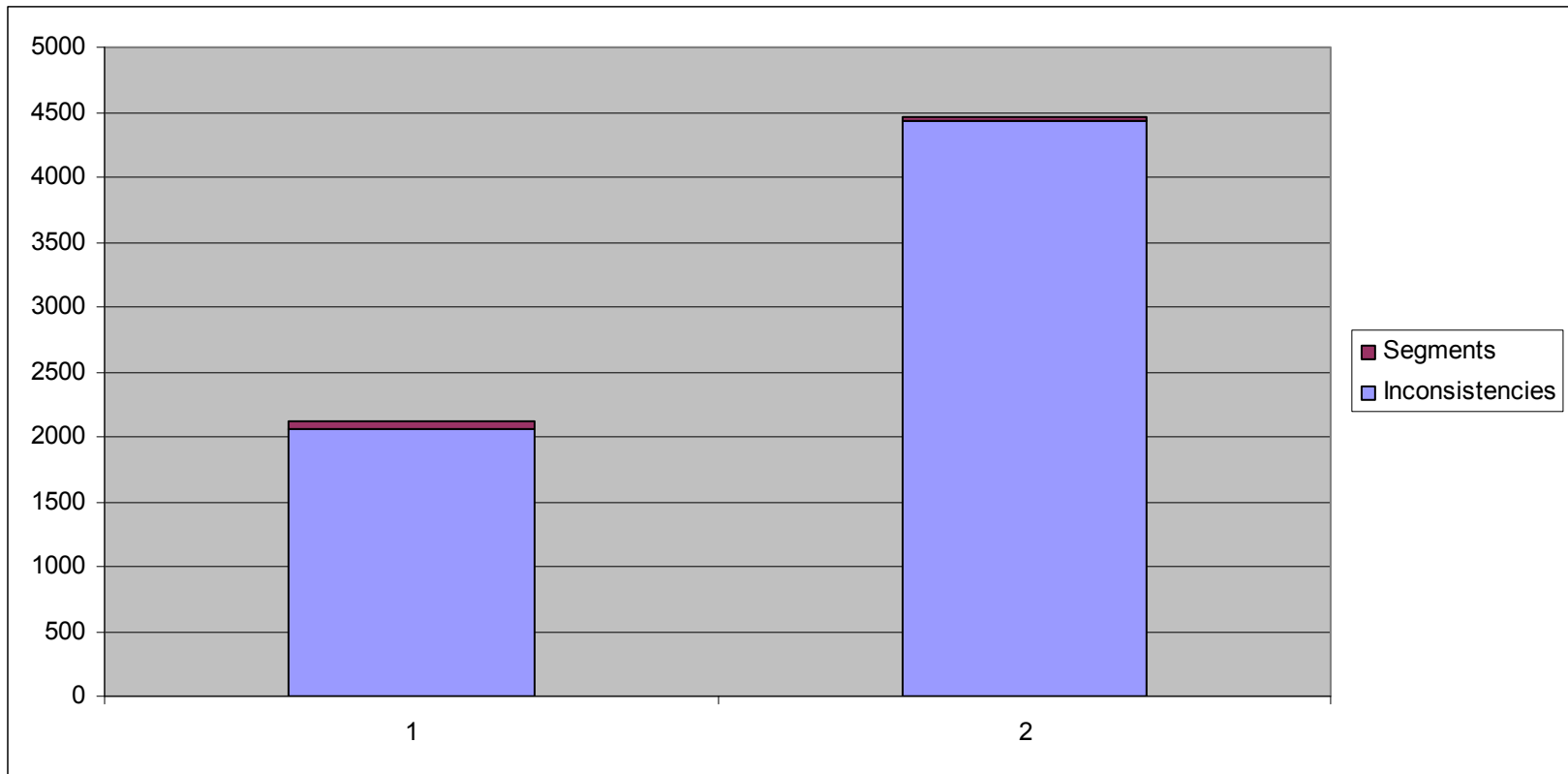
Lower Comprehension

Lower Productivity

Average Percentage of Inconsistencies



Frequency of Inconsistencies



Methodology

- Files presented in .tmw format →
SDL Trados 2007
- Exported as text files → Maruo text editor
- Grep command to separate by language
- US English tag: <seg L=EN_en>
- German tag: <seg L=DE_de>

Methodology (2)

- Import US_en files to Wordlist
- Search for 2-3 word clusters
- Aligned texts → Paraconc bilingual concordance
- Search for word clusters and sort
- Search for inconsistencies
- Categorise

are multiple icons in the file, the index number specifies the icon that appears. (30663): <Seg L=EN-US>For example, 1 means the first icon embedded in the program file. (30670): <Seg L=EN-US>{\cs6\fl\cf6\lang1024 <
office XP installation. (27665): <Seg L=EN-US>Run System Files Update from: (27672): <Seg L=EN-US>For example, 1033 for English. (27679): <Seg L=EN-US>Perform a compare scan and build a system update packag
\$eg L=EN-US>All positive integer values are seconds since January 1, 1970. (16191): <Seg L=EN-US>For example, 1069189856 is a time on November 19, 2003. (16198): <Seg L=EN-US>The null value is 1 / 1 / 1970.
DLLRETURN\$ variable uses an integer, unless you add str after the variable. (33198): <Seg L=EN-US>For example, 128, str is an acceptable variable name. (33205): <Seg L=EN-US>Specifying whether the .dll returns a
es it. (32109): <Seg L=EN-US>You can use a wildcard as the last character. (32119): <Seg L=EN-US>For example, (32126): <Seg L=EN-US>FILES=* (32133): <Seg L=EN-US>Installation searches for a text line that sta
33345): <Seg L=EN-US>There are nine variables, \$ASK1\$ through \$ASK9\$. (33358): <Seg L=EN-US>For example: (33365): <Seg L=EN-US>For example: (33372): <Seg L=EN-US>{\cs6\fl\cf6\lang1024 <:cs "user_en
: variables, \$ASK1\$ through \$ASK9\$. (33358): <Seg L=EN-US>For example: (33365): <Seg L=EN-US>For example: (33372): <Seg L=EN-US>{\cs6\fl\cf6\lang1024 <:cs "user_entry" 1}>ASK3: (33379): <Seg L=EN-US:
e this function to determine if a directory exists by using a \ after the directory name, ISFILE(c:\xyz\), for example. (34067): <Seg L=EN-US>{\cs6\fl\cf6\lang1024 <:mk 112<:crmk 112}>}If Locate() (34074): <Seg L=E
</cs>. (8236): <Seg L=EN-US>To restart the server, do one of the following: (8243): <Seg L=EN-US>For example (8250): <Seg L=EN-US>Creates an image file (8257): <Seg L=EN-US>Enter the full path of the executab
stored in the \$ASKn\$ variables are strings and are compared alphabetically. (33452): <Seg L=EN-US>For example, 99 is considered larger than 100 because the first character 9 is larger than the first character 1. (3345
n set up a group for the users and set the appropriate properties in the group. (13721): <Seg L=EN-US>For example, a group could correspond to a department of your organization, where users perform similar tasks with th
non-fixed disks (49697): <Seg L=EN-US>If files residing on non-fixed disks are included in a package, for example, a floppy drive or a network share, Client Migration has a safety feature that prevents these be
ng1024 <:v Migration Server" 3}>. (352): <Seg L=EN-US>specifying (359): <Seg L=EN-US>For example, adm-migrate:8080. (366): <Seg L=EN-US>{\cs6\fl\cf6\lang1024 <:imk 29}>}If the {\cs6\fl\cf6\lang10
Restored files inherit the permissions of the folder in which they are restored. (45795): <Seg L=EN-US>For example, all files that are restored to the User_1 profile inherit the security descriptors for the User_1 profile. (4580
<Seg L=EN-US>ID_COM_REGSRVR (32735): <Seg L=EN-US>Register a self-registering component, for example, an .ocx or .dll file. (32742): <Seg L=EN-US>{\cs6\fl\cf6\lang1024 <:mk 67<:crmk 67}>}Rename (327-
all from a CD-ROM or network drive, or to back up files on the user's system. (18940): <Seg L=EN-US>For example, c:\ installs the files to the root directory on drive C. (18947): <Seg L=EN-US>Use \ to install the files c
41): <Seg L=EN-US>In some cases the absolute path might not be appropriate. (548): <Seg L=EN-US>For example, C:*.ppt with 'Include Subfolders' selected is an absolute path. (555): <Seg L=EN-US>System-defined
/E\$ (35084): <Seg L=EN-US>\$DEFAULTDRIVE\$ is assigned the value of the default installation drive, for example, C. Diese Variable ist vergleichbar mit der Variablen \$DEFAULTDIR\$. (35091): <Seg L=EN-US>Howeve
able (19812): <Seg L=EN-US>\$DEFAULTDRIVE\$ is assigned the value of the default installation drive, for example, C. This variable is similar to \$DEFAULTDIR\$. (19819): <Seg L=EN-US>\$DLLRETURN\$ and \$DLLRET
ate the files with a pipe (|) delimiter. (18514): <Seg L=EN-US>Use /R to include subfolders in a path (for example, c:\docs*.doc/R). (18521): <Seg L=EN-US>This provides the same functionality as the Add Files dial
S) This variable is assigned the directory path of the Windows help directory. (35733): <Seg L=EN-US>For example, copying application help files with the FILE command into the Windows 95/NT help directory. (35740):
need to specify the LDAP name of the user container that you want to import. (16674): <Seg L=EN-US>For example, {\cs6\fl\cf6\lang1024 <:cs "user_entry" 2}>}LDAP: (16681): <Seg L=EN-US>\\myserver.com, CN =
=EN-US>from a domain (38734): <Seg L=EN-US>The results are as follows: (38743): <Seg L=EN-US>For example, {\cs6\fl\cf6\lang1024

} (38750): <Seg L=EN-US>On the Tools menu, click {\cs6\fl\cf6\l
package, then in the Extension box, type a wildcard character or parameters. (41633): <Seg L=EN-US>For example, *.doc to include only .doc files, or pr:* to include only files that begin with pr. (41640): <Seg L=EN-US>
ialog box in the {\cs6\fl\cf6\lang1024 <:v Migration Client" 1}>. (18531): <Seg L=EN-US>For example, /files: Āćāī-ĀĴC:\Documents*.doc|C:\My Music*.mp3Āćāī-ĀĴE (18538): <Seg L=EN-US>This inf
the overwrite parameter. (2068): <Seg L=EN-US>The default action is create. (2075): <Seg L=EN-US>For example, /files: Āćāī-ĀĴC:\Documents*.doc|C:\My Music*.mp3Āćāī-ĀĴE (2082): <Seg L=EN-US>This info
e {\cs6\fl\cf6\lang1024 <:fc 2> <:v Migration Server" 3> </fc>. (40586): <Seg L=EN-US>For example, http://adm-migrate (40596): <Seg L=EN-US>Logging on to the {\cs6\fl\cf6\lang1024 <:v

1 bedeutet beispielsweise, dass das erste Symbol angezeigt wird, das in die Programmdatei eingebettet ist. (30671):
(beispielsweise 1033 für Englisch). (27680):
Beispielsweise würde 1069189856 einen Zeitpunkt am 19. November 2003 stehen. (16199):
128, str ist beispielsweise ein gültiger Variablenname. (33206):
Beispiel: (32127):
Beispiel: (33366):
Beispiele: (33373):
Sie können mit dieser Funktion auch nachprüfen, ob ein Verzeichnis existiert, indem Sie nach dem Verzeichnisnamen ein \ setzen, z.B. ISFILE(c:\xyz\). (34068):
Beispiel: (8251):
Der Wert 99 gilt beispielsweise als größer, er als der Wert 100, da das erste Zeichen 9 größer ist als das erste Zeichen 1. (33460):
Eine Gruppe könnte beispielsweise einer Abteilung in Ihrem Unternehmen entsprechen, innerhalb derer die Benutzer ähnliche Aufgaben ausführen, die gleichen Anwendungen verwenden und ihre D
Wenn Dateien, die sich auf Wechseldatenträgern (z. B. auf einer Diskette oder einer Netzwerkfreigabe) befinden, in ein Paket aufgenommen werden, wird durch eine Sicherheitsfunktion von Client Migratio
Beispiel: adm-migrate:8080. (367):
Alle im Profil User_1 wiederhergestellten Dateien erhalten beispielsweise die Sicherheitsdeskriptoren für das Profil User_1. (45803):
Mit diesem Befehl können Sie eine sich selbst registrierende Komponente, wie z.B. eine .ocx- oder .dll-Datei, registrieren. (32743):
Beispielsweise werden durch Angabe von "C:\\" alle Dateien in das Stammverzeichnis von Laufwerk C: installiert. (18948):
Beispiel: C:*.ppt mit Auswahl von "Unterdorder einschließen" ist ein absoluter Pfad. (556):
\$DEFAULTDRIVE\$ wird der Wert des Standardinstallationslaufwerks (beispielsweise C) zugewiesen. (35092):
\$DEFAULTDRIVE\$ wird der Wert des Standardinstallationslaufwerks (beispielsweise C:) zugewiesen. Diese Variable ist vergleichbar mit der Variablen \$DEFAULTDIR\$. (19820):
Verwenden Sie /R, um Unterdorder eines Pfads mit einzuschließen (z.\cs6\fl\cf6\lang1024 <:hs>}B. C:\Dokumente*.doc/R). (18522):
Beispiel: Kopieren von Anwendungshilfedateien mit dem Befehl FILE in das Windows 95/NT-Hilfverzeichnis. (35741):
Ein Beispiel: {\cs6\fl\cf6\lang1024 <:cs "user_entry" 2}>}LDAP: (16682):
Ein Beispiel: {\cs6\fl\cf6\lang1024

} (38751):
Beispielsweise "*.doc", wenn Sie nur Doc-Dateien einschließen möchten, oder "pr.*", wenn Sie nur Dateien einschließen möchten, die mit "pr" beginnen. (41641):
Beispiel: /files: Āćāī-ĀĴC:\Documents*.doc|C:\My Music*.mp3Āćāī-ĀĴE (2082): <Seg L=EN-US>This info

Methodology (2)

- Import US_en files to Wordlist
- Search for 2-3 word clusters
- Aligned texts → Paraconc bilingual concordance
- Search for word clusters and sort
- Search for inconsistencies
- Categorise

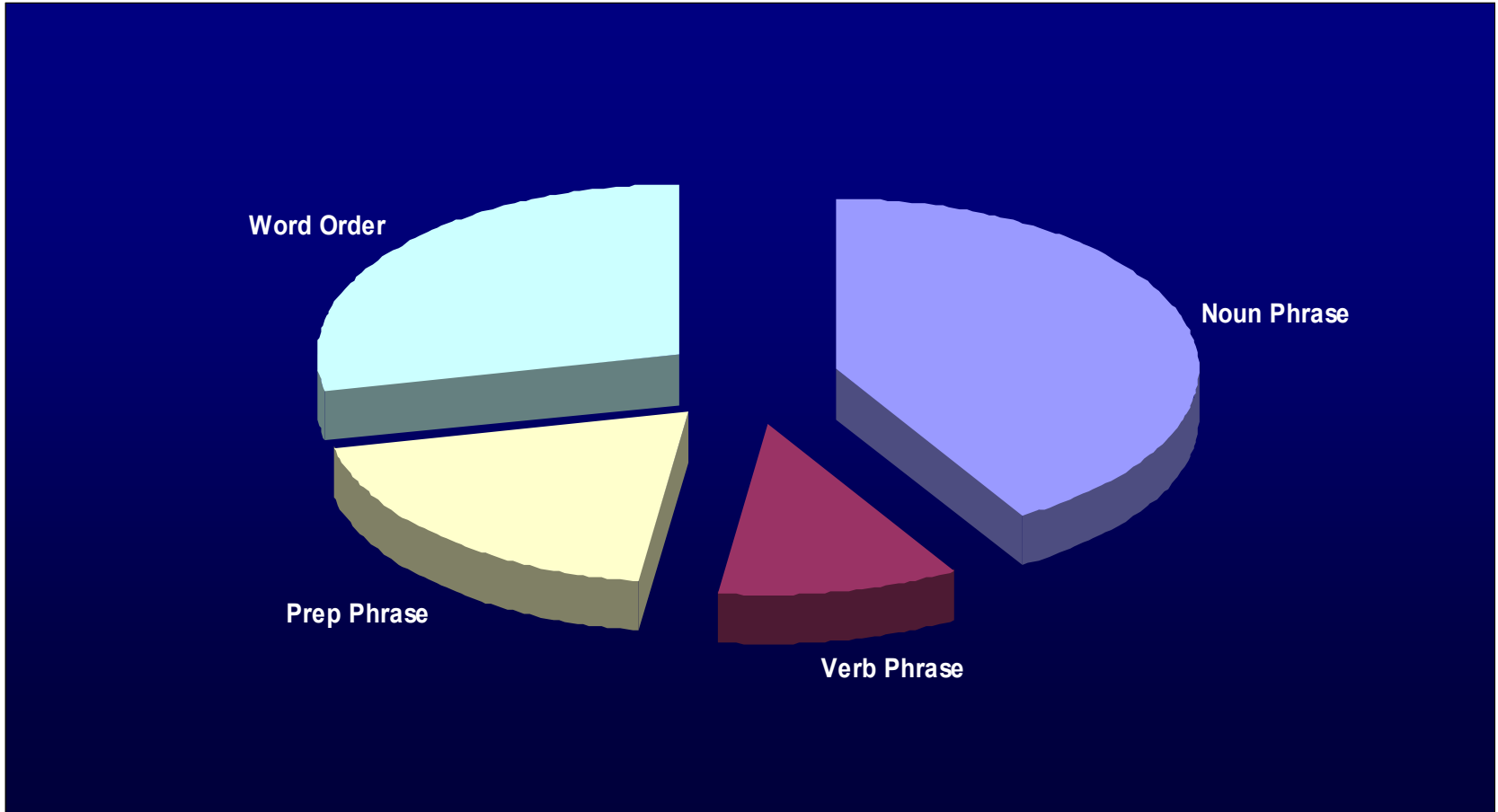
Categories: Noun Phrases

- <Seg L=EN-US>If the user selects the first and third groups (which **the administrator** assigned to group numbers 10 and 30), then IF GROUP = 10,20 is true because group number 10 is selected.
- <Seg L=DE-DE>Wenn der Benutzer die erste und dritte Gruppe auswählt (die **der Administrator** den Gruppennummern 10 und 30 zugeordnet hat), dann ist IF GROUP = 10,20 wahr, da die Gruppennummer 10 ausgewählt ist.
- <Seg L=DE-DE>Wenn der Benutzer die erste und dritte Gruppe auswählt (die **der Verwalter** den Gruppennummern 10 und 30 zugeordnet hat), dann ist IF GROUP = 10,20 wahr, da die Gruppennummer 10 ausgewählt ist.

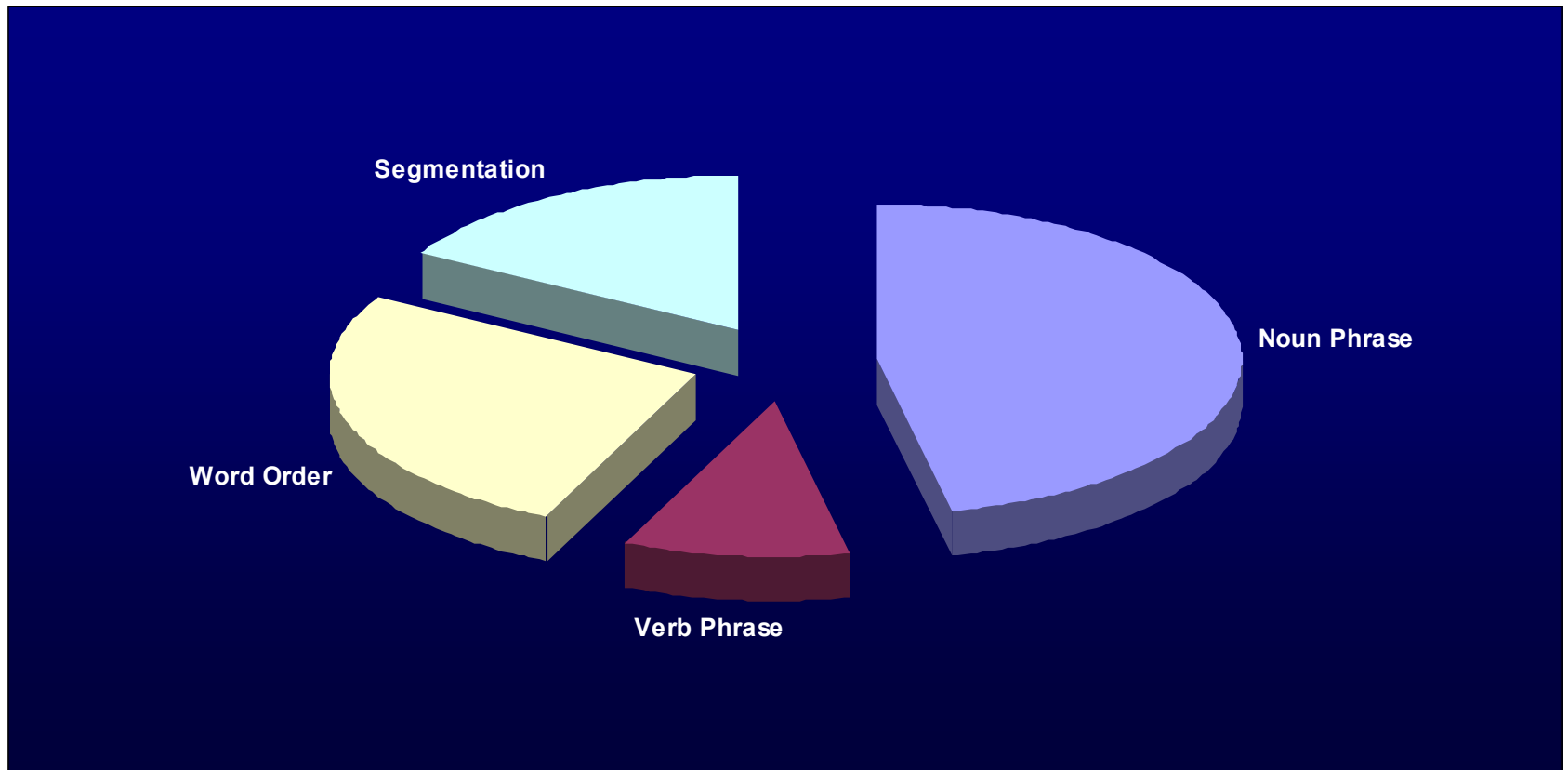
Categories: Word Order

- <Seg L=EN-US>The selected user profile is restored to **the user who is currently logged on.**
- <Seg L=DE-DE>Das ausgewählte Benutzerprofil wird für **den derzeit angemeldeten Benutzer** wiederhergestellt.
- <Seg L=DE-DE>Das ausgewählte Benutzerprofil wird für **den Benutzer** wiederhergestellt, **der derzeit angemeldet ist.**

2004 Distribution of Inconsistencies



2008 Distribution of Inconsistencies



Categories: Segmentation

- **<Seg L=EN-US> This is updated at the end of the [Product] process, which might require you to reinsert the first disk in the span set.** [Product name] prompts you for the first disk in the span set and for subsequent volumes when restoring...
- **<Seg L=EN-US> This is updated at the end of the [Product] process, which might require you to reinsert the first disk in the span set.**
- **<Seg L=DE-DE> Diese Informationen werden am Ende des Duplizierungsprozesses aktualisiert. Für die Aktualisierung ist es notwendig, dass Sie den ersten Datenträger des verteilten Datenträgersatzes erneut einlegen.** Beim Wiederherstellen aus einer Image-Datei...
- **<Seg L=DE-DE> Diese Informationen werden am Ende des [Product]-Prozesses aktualisiert, möglicherweise müssen Sie dann den ersten Datenträger des gesplitteten Satzes nochmals einlegen.**

Conclusions & Recommendations

- Controlled Language, CMS →
Fewer inconsistencies
- Noun Phrases: Stylistic choice or workflow issues?
- Tighter control on terminology vs glossaries
- Consistent segmentation rules
- Maintaining TM database is important
- Future work: Expand study, Suggest improvements

Thank You!

Email: jossmo@gmail.com
Research funded by SFI/CNGL