

Ain Shams University
Faculty of Al-Asun
Department of English



**Cohesion in Arabic into English Translation:
A Corpus-Based Investigation of the Translation Universals**

**A Thesis in Partial Fulfillment of the Requirements
of the Master's Degree of Arts**

Submitted by
Yasmin Ossama Al-Kashef

Under the supervision of	
Dr. Mona Fouad Attia	Dr. Amal Mohammad Abdel Maqsoud
Associate Professor of Linguistics	Lecturer of Translation & Linguistics
English Department	English Department
Faculty of Arts	Faculty of Al-Asun
Helwan University	Ain Shams University

2011



جامعة عين شمس

كلية الألسن

قسم اللغة الإنجليزية

التماسك اللغوي في الترجمة من العربية إلى الإنجليزية:

دراسة نصية في السمات العامة للترجمة

رسالة لنيل درجة الماجستير

مقدمة من ياسمين أسامة الكاشف

تحت إشراف

د/ أمل محمد عبد المقصود

مدرس اللغويات والترجمة

قسم اللغة الإنجليزية

كلية الألسن

جامعة عين شمس

أ.م. د/ منى فؤاد عطية

أستاذ مساعد اللغويات

قسم اللغة الإنجليزية

كلية الآداب

جامعة حلوان

جامعة عين شمس، كلية الألسن، 2011

Abstract

Al-Kashef, Yasmin Osama. *Cohesion in Arabic into English Translation: A Corpus-Based Investigation of the Translation Universals*. MA. Thesis, Faculty of Al-Asun (Languages). Ain Shams University, 2011.

The study sets out to investigate the translation universals, namely explicitation, simplification and normalization in translated texts through the use of comparable corpus comprising two sub-corpora: English news stories translated from Arabic and original non-translated English news stories. Recent literature has identified such features as simplification, explicitation and normalization in translated texts into major Indo-European languages, and translation researchers are keen to know if these phenomena are also present in non-European languages like Arabic. By applying corpus-based methodology, the study focuses on cohesion in English and Arabic, in particular the frequency and usage of the different cohesive devices in translated texts compared to non-translated English. The one million-word corpus of news stories translated from Arabic into English is part of the Linguistic Data Consortium (LDC) parallel corpus of news stories from the *Ummah* Press Service, whereas the other sub-corpus comprises one million words from the English Treebank corpus comprising news stories from the *Wall Street Journal*. The results are further tested using a sample of translated texts from the LDC corpus together with their original texts in order to verify the observations of studying the comparable corpus. The study concludes by affirming the hypothesis but with some reservations on the general statements made about the translation universals.

Keywords:

Translation Universals, Corpus-based translation studies, Descriptive translation Studies, Discourse analysis, Cohesion

List of Tables

Table 1	The use of "same" as an accented form for the different types of substitution	39
Table 2	Categories of additive conjunction	45
Table 3	Categories of adversative conjunction	46
Table 4	Categories of causal conjunction	47
Table 5	Categories of causal conjunction	48
Table 6	The main features of the comparable corpus	90
Table 7	The first ten lines of the TC histogram	94
Table 8	Reference items investigated in the present study	95
Table 9	Substitution items investigated in the present study	96
Table 10	Ellipsis items investigated in the present study	97
Table 11	Conjunction items investigated in the present study	97
Table 12	Type-token ratio in TC and NTC	102
Table 13	Ellipsis items investigated in the present study	116
Table 14	Reference in TC and NTC	128
Table 15	Personal reference in TC and NTC	134
Table 16	Substitution in TC and NTC	153
Table 17	Collocate terms in TC and NTC	183

List of Figures

Figure 1	The map of translation studies as envisaged by Holmes	10
Figure 2	Taxonomic relations tree diagram	51
Figure 3	KWIC window	78
Figure 4	A snapshot of the SCP output file	92
Figure 5	Most common collocate terms in TC as identified by NSP	93
Figure 6	Ellipsis in TC and NTC	117
Figure 7	Conjunction in TC and NTC	119
Figure 8	Causal conjunction in TC and NTC	120
Figure 9	1 st , 2 nd and third person pronouns in TC and NTC	134
Figure 10	Demonstrative reference in TC and NTC	142
Figure 11	Comparative reference in TC and NTC	151
Figure 12	Nominal substitution in TC and NTC	154
Figure 13	Clausal substitution in TC and NTC	154
Figure 14	Conjunctions in TC and NTC	163
Figure 15	Additive conjunctions in TC and NTC	164
Figure 16	Simple and complex additive conjunctions in TC and NTC	165
Figure 17	Adversative conjunctions in TC and NTC	171
Figure 18	Simple and emphatic adversative conjunctions in TC and NTC	172
Figure 19	Causal conjunctions in TC and NTC	179
Figure 20	Collocate terms in TC and NTC	182

Symbols used in the Phonemic Transcription of Arabic Forms

A) Consonants

/ʔ/:	glottal plosive /ء/
/b/:	voiced bilabial plosive /ب/
/t/:	non-emphatic voiceless denti-alveolar plosive /ت/
/θ/:	voiceless dental fricative /ث/
/j/:	voiced velar fricative /ج/
/ħ/:	voiceless pharyngeal fricative /ح/
/x/:	voiceless uvular fricative /خ/
/d/:	non-emphatic voiced denti-alveolar plosive /د/
/ð/:	non-emphatic voiced dental fricative /ذ/
/r/:	Voiced alveolar trill /ر/
/z/:	Voiced alveolar fricative /ز/
/s/:	Non-emphatic voiceless alveolar fricative /س/
/ʃ/:	Voiceless palato-alveolar fricative /ش/
/S/:	Emphatic voiceless alveolar fricative /ص/
/D/:	Emphatic voiced denti-alveolar plosive /ض/
/T/:	Emphatic voiceless denti-alveolar plosive /ط/
/ð/:	Emphatic voiced dental fricative /ظ/
/ʕ/:	Voiced pharyngeal fricative /ع/
/ʁ/:	Voiced uvular fricative /غ/
/f/:	Voiceless labio-dental fricative /ف/
/q/:	Voiceless uvular plosive /ق/
/k/:	Voiceless velar plosive /ك/
/l/:	Voiced denti-alveolar lateral /ل/
/m/:	Voiced bilabial /م/
/n/:	Voiced denti-alveolar nasal /ن/
/h/:	Voiceless glottal fricative /ه/
/w/:	Labio-velar semi-vowel /و/
/y/:	Voiced palatal semi-vowel /ي/

B) Vowels

/a/:	Front open vowel (short & long) (فتحة)
------	--

/i:/ Front close vowel (short & long) (كسرة)

/u:/ Back close vowel (short & long) (ضمة)

Notes

1. Length of vowels is indicated by doubling the symbol.
2. Geminated Consonants are indicated by doubling the symbol.

(Ezzat, 1975)

List of Abbreviations

BNC	British National Corpus
Cobuild	Collins Birmingham University International Language Database
DTS	Descriptive Translation Studies
EAGLES	Expert Advisory Group on Language Engineering Standards
ECC	English Comparable Corpus
GUI	Graphical User Interface
ICAME	International Computer Archive of Modern and Medieval English
LDC	Linguistic Data Consortium
LOB	Lancaster-Oslo-Bergen Corpus
NSP	N-gram Statistics Package
NTC	Non-translated Corpus (corpus made of texts originally written in English)
OC	Original Corpus (texts originally written in Arabic)
SCP	Simple Concordance Program
SL	Source language
ST	Source Text
TC	Translated Corpus
TL	Target language
TT	Target Text
UIH	Unique Items Hypothesis

Table of Contents

Acknowledgements	i
Abstract	ii
List of abbreviations	iii
List of Tables	iv
List of Figures	v
Symbols used in the Phonemic Transcription of Arabic Forms	vi
Table of Contents	viii
Introduction	1
Part I: theoretical Preliminaries	7
Chapter 1: Translation Universals and Cohesion	8
1.1 Translation Universals	9
1.1.1 Descriptive Translation Studies	9
1.1.2 The Search for Regularities	12
1.1.3 Plethora of terms	12
1.1.4 The Three Universals	15
1.1.4.1 Explicitation	16
1.1.4.2 Simplification	20
1.1.4.3 Normalization	22
1.1.5 Other Universals	24
1.1.6 Different Typology	27
1.1.7 The Universals: A Debatable Concept	28
1.2 Cohesion	30
1.2.1 Types of cohesive devices	32
1.2.1.1 Grammatical Cohesion	32
1.2.1.1.1 Reference	32
1.2.1.1.2 Substitution	37
1.2.1.1.3 Ellipsis	40
1.2.1.1.4 Conjunctions	42
1.2.1.2 Lexical cohesion	49
1.2.1.2.1 Reiteration	50
1.2.1.2.2 Collocation	53
1.2.2 Cohesion in Arabic	55
1.2.2.1 Early discussions of cohesion in Arabic	55
1.2.2.2 Modern discussions of cohesion in Arabic	56
1.2.3 Cohesion and Translation	61
1.3 Conclusion	62
Chapter 2: Corpus Linguistics and Corpus-Based Translation Studies	63
2.1 Corpus	64
2.2 Corpus Linguistics	65
2.3 Corpus Linguistics: A Historical Overview	66
2.4 The Development of Corpora	68
2.4.1 First generation corpora	68
2.4.2 Second Generation Corpora	69

2.5 Corpus Types	70
2.6 Corpus Design	71
2.6.1 Size	72
2.6.2 Representativeness	72
2.6.3 Balance	73
2.7 Corpus Pre-processing	74
2.8 Corpus Analysis Tools	75
2.8.1 Word frequency lists	76
2.8.2 Concordancer	77
2.8.3 Collocation generator/extractor	78
2.9 Corpus-Based Translation Studies	79
2.10 Comparable Vs. Parallel Corpora	81
2.10.1 Parallel Corpus	82
2.10.2 Comparable Corpus	84
2.11 Conclusion	86
Chapter 3: Methodology	88
3.1 Corpus	89
3.1.1 A corpus of translated English (TC)	89
3.1.2 A corpus of original English (NTC)	90
3.2 Tools	90
3.2.1 Simple Concordance Program (SCP)	90
3.2.2 Ngram Statistics Package (NSP)	91
3.3 Procedures	91
3.4 Limitations and controls	95
3.4.1 Reference	95
3.4.2 Substitution and Ellipsis	96
3.4.3 Conjunctions	97
3.4.4 Lexical cohesion	98
3.5 Conclusion	98
Part II: Application	99
Chapter 4: Explication	100
4.1 Reiteration	101
4.2 Ellipsis	116
4.3 Conjunction	118
4.3.1 Causal Conjunctions	119
4.4 Reference	125
4.5 Conclusion	128
Chapter 5: Simplification	131
5.1 Reference	132
5.1.1 Personal Reference	133
5.1.2 Demonstrative Reference	141
5.1.3 Comparative Reference	151
5.2 Substitution	153
5.3 Conjunctions	157
5.4 Conclusion	160

Chapter 6: Normalization	162
6.1 Conjunctions	163
6.1.1 Additive Conjunctions	164
6.1.2 Adversative Conjunctions	171
6.1.3 Causal Conjunctions	179
6.2 Collocation	180
6.3 Conclusion	187
Chapter 7: Conclusion	189
7.1 Research Steps	190
7.2 Problems Encountered during the Study	190
7.3 Results	191
7.3.1 Grammatical Cohesion	192
7.3.1.1 Reference	192
7.3.1.2 Substitution	192
7.3.1.3 Ellipsis	192
7.3.1.4 Conjunction	193
7.3.2 Lexical Cohesion	193
7.3.2.1 Reiteration	193
7.3.2.2 Collocation	194
7.4 Overall Conclusions	194
7.4.1 Explication	194
7.4.2 Simplification	194
7.4.3 Normalization	195
7.5 Recommendations	195
Bibliography	197

