Μετα

Target Text Contraction in English-into-Korean Translations: A Contradiction of Presumed Translation Universals?

Ho-Jeong Cheong

Citer cet article

Résumé de l'article

Cet article contredit l'hypothèse qui a cours parmi les partisans des universels de traduction selon laquelle l'explicitation, procédé consistant à expliquer clairement les choses plutôt que de les laisser implicites dans le texte traduit, est un universel de traduction (UT) potentiel quelles que soient les deux langues présentes dans le processus de traduction. Au moyen de l'étude d'un nouveau corpus parallèle de 517 609 mots, on a notamment observé aussi bien l'implicitation et la contraction du texte d'arrivée qui l'accompagne que l'explicitation et l'expansion du texte traduit que cela implique dans des traductions entre le coréen et l'anglais. On a constaté que la direction des combinaisons linguistiques dans les traductions employant les mêmes langues est significative et on a introduit et vérifié la validité des quatre unités de mesure conçues pour que cette étude saisisse les différents aspects de l'explicitation/implicitation et de l'expansion/contraction du texte traduit qui en découlent.

Target Text Contraction in English-into-Korean Translations: A Contradiction of Presumed Translation Universals?*

HO-JEONG CHEONG
Hankuk University of Foreign Studies, Seoul, Korea
angelhjchg@hufs.ac.kr

RÉSUMÉ

Cet article contredit l'hypothèse qui a cours parmi les partisans des universels de traduction selon laquelle l'explicitation, procédé consistant à expliquer clairement les choses plutôt que de les laisser implicites dans le texte traduit, est un universel de traduction (UT) potentiel quelles que soient les deux langues présentes dans le processus de traduction. Au moyen de l'étude d'un nouveau corpus parallèle de 517 609 mots, on a notamment observé aussi bien l'implicitation et la contraction du texte d'arrivée qui l'accompagne que l'explicitation et l'expansion du texte traduit que cela implique dans des traductions entre le coréen et l'anglais. On a constaté que la direction des combinaisons linguistiques dans les traductions employant les mêmes langues est significative et on a introduit et vérifié la validité des quatre unités de mesure conçues pour que cette étude saisisse les différents aspects de l'explicitation/implicitation et de l'expansion/contraction du texte traduit qui en découlent.

ABSTRACT

This paper contradicts the prevailing assumptions among the advocates of translation universals (TU's) that explicitation, a translation behavior which consists of spelling things out rather than leaving them implicit in translation, is a potential TU, irrespective of the specific language pairs involved in the process of translation. Specifically, via a study employing a newly built 517,609-word parallel corpus, it is shown that implicitation and the subsequent TT contraction as well as explicitation and TT expansion entailed were both observed in translations involving Korean and English. The significance of the direction of language combinations in translations employing the same language pair was identified, together with the introduction and verification of the validity of the four measurement units devised for this study to capture diverse aspects of explicitation/implicitation which in turn entail TT expansion/contraction.

초록

본 논문은 원문 텍스트에 암묵적으로 나타나있는 의미를 '번역 과정에서 언어적으로 구현하여 의미의 명확성을 제고하는 번역 현상을 가리키는 '외연화'가 번역언어에 상관없이 나타나는 '번역 보편소' 후보라는 '번역 보편소' 주창자들의 가설을 실증적 자료분석을 통해 반박한다. 구체적으로는 한국어와 영어 간의 번역 텍스트에서는 외연화와 그에 따른 번역텍스트 확장현상 이외에 원문 텍스트에 명시적으로 구현되어 있는 의미를 암묵적 추론이 가능한 방식으로 구성하는 '내포화'와 그로 인한 번역텍스트 축소현상도 관찰되고 있음을 보고한다. 연구방법론으로 본 연구를 위해 새로 구성된 517,609 단어(토큰 기준)의 병렬 코퍼스를 다양한 통계기법을 사용, 분석한다. 이를 통해 동일 언어 쌍이 번역언어로 사용되는 경우에도 언어별 번역방향(한영 및 영한)이 연구대상인 번역현상에 중요한 차이를 가져온다는 사실을 보고하는 한편, 외연화/암묵화와 그로 인한 번역 텍스트 확장/축소의 다양한 측면을 포착하기 위해 고안된 네 가지 측정단위의 타당성을 입증한다.

I. Introduction

'Target text (TT) contraction' signifies the tendency of a translated text to become shortened in relation to the source text (ST). The attention of Translation Studies (TS) scholars who have entertained the idea of 'translation universals' (TU's), or the assumption of certain attributes which typically recur only in TT's in comparison with ST's, has thus far almost exclusively been devoted to the phenomenon of explicitation, making it the strongest possible candidate TU (Baker 1993; Vanderauwera 1985; Leppihalme 1997; Øverås 1998; Toury 1995). The simplest definition of explicitation is that it is an "overall [translation behavior] to spell things out, rather than leaving them implicit in translation" (Klaudy 1998: 80), and probably the most comprehensive definition comes from Vanderauwera (1985), who proposes the following list of explicitation repertoires: expansion of condensed passage; addition of modifiers, qualifiers and conjunctions to achieve greater transparency; and addition of extra information and insertion of explanations, among many others. A review of various definitions of the term leads to the conclusion that the most obvious text-surface manifestation of explicitation is TT expansion, or the TT behavior of lengthening in relation to a ST as the result of addition/insertion of explanatory information. On the other hand, TT contraction is the text-surface realization of 'implicitation.' If not entirely neglected in the literature (Nord 1991; Leppihalme 1997), the relative neglect of 'implicitation,' and therefore of TT contraction, in TS can be sensed from the fact that few scholars have given specific definitions of the latter concept, feeling it suffices to simply add "or vice versa" at the end of the definition of explicitation. Only Øverås explicitly defines implicitation as "instances where explicit ST items are rendered by ambiguous [or even zero (ø)] TT items, but where recoverability in the immediate TT environment makes [the items still] retrievable" (1998: para. 9). As a result, only scant attention has thus far been paid to TT contraction.

Whether referring to explicitation or implicitation, or to TT expansion or contraction as their respective realizations, one thing that should be noted is that TU's are only a hypothetical concept. Toury's (1995) proposal for experimental studies of the correlation between explicitness, or the level of explicitation, and readability with a view to assessing the varying extent to which explicitation may be applied was made based on the assumption that explicitation is an obvious fact. Baker, who pointed to explicitation as one of the four potential TU's, observed that "[TU's] are thought to be independent of the influence of the specific language pairs involved in the process of translation" (1993: 243). However, whether explicitation and subsequent TT expansion are observable or even dominant in all translations irrespective of the working language pairs involved should be tested against empirical data before we grant them the status of a TU.

The present study is a response to this challenge. Research has been devised with the following aims:

(1) To test whether TT expansion rather than TT contraction is an observable or even dominant phenomenon in translations between the Korean and English languages

(2) If so, to test whether the direction of translating language combination has any influence upon the translation behavior in question, and

(3) To identify 'TT expansion/contraction factors,' or factors affecting the phenomenon in question.

In designing a research plan to satisfy the above-mentioned aims, three methodological decisions were made. First, a corpus-based approach was adopted and an independent parallel corpus[1] was compiled which comprises 517,609 token words (299,630 in English and 218,979 in Korean) in 322,945 sentences in 316 authentic texts. Second, four different measurement units were devised in order to capture diverse aspects of the translation phenomenon in question. Third, the statistical analyses of the corpus were made at the text level instead of at either word or sentence levels.

II. Theoretical Framework

2.1. Explicitation/Implicitation

2.1.1 Putting the Concepts into Perspective

A look at the following examples facilitates our understanding of explicitation and implicitation.

[1]

ST1: 이 그림은[2]   고려 말   삼은   가운데 한   분이었던
 [*i geurim-eun   goryeo-mal sam-eun   gaunde han   bun-i-eoss-deon*]
 'this painting-SBJ[3] Goryo-end three-*eun*   among one   person-HNR-being-PST-ABJ-END'
 목은 이색 (1328~1396)   선생의   영정으로 1884년   (헌종 10)에
 [*Mogeun Yi Saek (1328~1396) seonsaeng-eui yeongjeong-euro 1884-nyeon (Heonjeong 10)-e*]
 'Mogeun Yi Saek (1328~1396) master-PSS portrait-being 1884-year (King *Heonjeong* 10)-in'
 그려졌다
 [*geuryeo-ji-eoss-da*]
 'paint-PSS-PST-END'

TT1: It is the portrait of M**ogeun** (**Yi Saek's** penname) **Yi Saek** (1328-1396) drawn in 1384 (10th year of King **Heonjong's reign**). Yi was one of the three greatest scholars of the **Goryo** Dynasty who happened to share the '**-eun'** syllable in their respective pennames.[4]

[2]

ST2: **In this tiny peninsular country, located at the northeastern tip of the Asian continent** with a long history and many traditions, insurance contract practice is also closely related to local conventions

TT2: 유구한   역사와   전통을   자랑하는   한국에서는 보험
 [*yuku-han   yeoksa-wa   jeontong-eul   jarang-ha-neun   hanguk-eseoneun boheom*]
 'time-honoredness-ADJ HISTORY-CNN. TRADITION-OBJ PRIDE-VRB-ADJ KOREA-SBJ INSURANCE'
 계약   관행   역시 사회적   관습과   밀접한   연관을   갖는다
 [*kyeyak gwanhaeng yeoksi sahoe-jeok gwanseup-gwa miljeop-han yeongwan-eul gan-neun-da*]
 'contract convention too society-ADJ PRACTICE-CNN CLOSENESS-ADJ RELATION-OBJ HAVE-PRT-END'

In example [1], '*Yi Saek*' is a well-established scholar who lived towards the end of the *Goryo* Dynasty of Korea. Yi Saek was also known as '*Mogeun,*' which was an appellation similar to a penname used due to a then-prevalent custom in Korea not to call

adults by their names. 'Sam-eun,' i.e., the three eun's, is a kind of a collective term referring to the three most authoritative scholars who exerted their influences concurrently and happened to share the syllable 'eun' in their respective pennames. The target readership of the TT would be oblivious to these references unless some explanatory information is supplied to supplement what is provided in the ST. The motivation for supplying such information is to make the corresponding ST segments explicit enough to be accessible to the target readership. The assumed target readership in this case is speakers of English, who generally do not have much background knowledge of Korea or of its traditions and linguistic conventions. Keeping in mind the definition of explicitation as a tendency to make the implicit ST segment explicit by, among other things, adding explanatory information, a striking example of explicitation is offered here by the occurrence of 'Sam-eun,' a single-word ST segment to which lavish explanatory information is supplied in the TT: 'the three greatest scholars of the Goryo Dynasty who happened to share the '-eun' syllable in their respective pennames.'

In example [2], the detailed description of the geographical location of Korea in the ST, bold-faced, is omitted in the TT since the translator appears to have assumed that the target readership of the TT, in this case Koreans, would easily retrieve that particular piece of information without explicit coding. Recall that the definition of implicitation includes "instances where explicit ST items are rendered by ambiguous [or even zero (ø)] TT items, but where recoverability in the immediate TT environment makes [the items still] retrievable" (Øverås 1998: para. 9). Example [2] is a clear instance of implicitation as the information regarding the geographical location of the country is substituted with a zero TT item. The TT readers, who are quite familiar with their mother country's geographical location, have no difficulty retrieving the corresponding information.

As can be readily observed from examples [1] and [2], the result of explicitation is TT expansion, and that of implicitation is TT contraction.

2.1.2. Motivations for Explicitation/Implicitation[5]

Three different theories can be drawn on to account for the motivation of explicitation/ implicitation. Firstly, Nida and Taber note the translation behavior of TT's of "amplification of information in the TT from implicit to explicit status" (1969: 164-65). In order to account for amplification, or the equivalent of explicitation, Nida resorts to the figurative 'channel model' and the concept of "information load" (1964: 131). According to the model, any well-constructed text is designed to fit the channel capacity of the language it employs. In translation, however, the capacity of the target language (TL) channel is narrower than that of the source language (SL). The information load, or the amount of information the reader has to possess in reading a text or a given segment thereof, is much higher for the TT readers than for the ST readers since the ST unfolds a text world that is unfamiliar to the readership of the TT. Due to this, immediate intelligibility of the ST is lower for the TT readers than for the ST readers. Acutely aware of this problem and perceiving a need to make necessary mediations, the translator supplies explanatory information in rendering the TT. That is, the translator exploits the mediation device of explicitation. Consequently, the TT becomes amplified, or longer than the ST.

According to the frame-and-scene model proposed by Fillmore (1982, 1985, 1992) and Tannen (1993a, 1993b), meaning cannot simply be conceived as a set of abstract semantic or situational features but as a holistic experiential knowledge which can be analyzed as consisting of frames and scenes. To be brief, a 'frame' signifies a particular organization of knowledge that stands as a prerequisite to our ability to understand the meaning of associated words and a 'scene' is an idealized schematized temporal sequence of doing things that share the same event or situation. A reader's ability to understand a text hinges on his sharing frames-and-scenes with the author of the text. The specific implications of this insight for translation come from Fillmore's categorization of frames-and-scenes: innate and learned. Unlike innate frames, learned frames vary, depending upon the language community in which a language user acquires a particular set of learned frames. The resulting culture-specificity of frames-and-scenes may cause problems in inter-cultural communications, including translation, when the interlocutors do not share the same frames and scenes. Having a bicultural command of the frames-and-scenes of the two language communities involved, and aware of its relevance, a competent translator can bridge gaps left due to the frame-and-scene disparities. Explicitation in the form of supplying bridging information is frequently used for this purpose, which in turn results in TT expansion.

Gutt (1990) advocates a more translation-oriented account by expanding on Sperber and Wilson's (1986) Relevance Theory. He argues that construction of any rational text is governed by the principle of communication: to verbalize 'only what is necessary' in such a way as to make the text as economical as possible as long as it is accessible to the readership. The need to make the communicative intention accessible to the intended reader (on the part of the text author), and to make the retrieval of the corresponding message as economical as possible (on the part of the text reader) results in texts comprising two dimensions of meaning: 'explicature' as explicitly verbalized meaning and 'implicature' as not explicitly verbalized, or implied, meaning. Few problems of interpretation arise when communications occur among members of the same cultural community, thanks to their sharing of both the 'cognitive environment,' understood as the totality of the extralinguistic encyclopedic knowledge, and the 'contextual assumptions' as "assumptions that language users entertain vis-à-vis the world" (Hatim 2001: 37). When communication occurs between members of different language communities, however, texts can be much less accessible due to possible "contextual gaps" (Gutt 1992: 68), or gaps existing between the contextual assumptions presupposed of the ST and TT readers. When contextual gaps occur, the translator will resort to explicating the implicit information of the ST and will provide any cultural information which cannot be assumed to be familiar to the TT readership. This quite naturally results in explanatory information being added to what is already in the ST, leading to TT expansion.

Compared to the theoretical treatment in the literature, as summarized above, however, empirically-oriented research on the topic of explicitation has not been widely undertaken. Blum-Kulka (1986) proposed the 'explicitation hypothesis,' referring to a general tendency toward increases in the explicitness level of translated texts, which she posits may be a universal strategy inherent in any process of language mediation. The list of empirical research efforts includes Seguinot (1988), who found that there was greater explicitness in translations in both French translations of English

ST's and English translations of French ST's as a result of improved topic-comment links, addition of linking words, and raising subordinate information into coordinate information. Olohan and Baker (2000) also identified a substantially heavier use of optional reporting *that* as a case of explicitation when compared with non-translational texts.

It is clear by now that (1) explicitation enjoys a privileged status, when compared with implicitation, and (2) explicitation/implicitation has been examined with a limited focus on some textual characteristics such as cohesion. However, as example [2] clearly demonstrates, implicitation is also attested in TT's, and explicitation/implicitation and TT expansion/contraction as their respective surface realizations should be approached at the text level as well as at the lower level of its constituents. Driven by the conviction that any system must have balance, the present study pursues implicitation as well as explicitation, that is, an investigation of TT contraction as well as TT expansion.

2.1.3. Diverse Aspects of TT Expansion/Contraction

What is required before deciding on measurement units with which to track TT expansion/contraction is to understand that TT expansion/contraction is a complex phenomenon comprised of diverse aspects.

Let us begin with example [3], which clearly indicates a shorter TT in comparison to its ST, i.e., a case of TT contraction measured in word counts, with 11 words used in the TT, compared to 18 words in the ST.

[3]
ST: The General Assembly shall be entitled to cancel the Membership by a two-thirds majority when a member has:
TT: 다음   경우   총회는   3분의 2   다수결로   회원 자격을
      [*Taeum gyeongu chonghoe-neun   sambun-ui-i   dasugyeol-lo   hoewon- jagyeog-eul*]
      'next occasion general assembly-sᴊʙ two-thirds   majority–through   membership–oʙᴊ'

        취소 할 수 있다.[6]
        [*Chwiso- hal   su   issda*]
      'canceling-ᴠʀʙ doing can'

On the other hand, in example [4], the sentence-initial connective *ttohan* 'and/again' of the ST, bold-faced and italicized, is omitted in the TT, which contradicts the existing report in the literature that sentence linking conjunctions are oftentimes inserted rather than omitted (Baker 1993; Leppihalme 1997; Vanderauwera 1985).

[4]
ST: 노동   신문 사설 게재 통해 郡의 역할 강화를   통한
      [*Nodong sinmun saseol gyejae tonghae gun-ui yeokal ganghwa- reul tonghan*]
      'Labor Daily editorial publishing through country-ᴘss role strengthening-oᴊʙ through'
        지방 경제   활성화   모색. *또한* 인민 경제   정보화를   촉구
      [*jabang gyeongje hwalseonghwa mosaek.* **ttohan** *inmin gyeongje jeongbohwa-reul chokku*]
      'local economy energization seeking   **also** people economy informatization- oʙᴊ urging'
TT: North Korea stressed the role of counties in invigorating regional economies in an editorial of the *Rodong Sinmun*, the official newspaper of the Korea Worker's Party. The same editorial **also** urged the implementation of the informatization of the people's economy.[7]

Example [5] is an instance of addition/insertion of extra explanatory information. Instead of stopping at providing phonetic representations of the culture-laden items, or the 'culturemes' (Vermeer 1989), all appearing in bold and italics, extensive supplementary information is added to the TT that is not originally found in the ST. For instance, in rendering *sarangchae*, extra information explaining the cultureme – [an area] used as a reception room for male guests – followed the term's Romanization, which only represents its phonetic values. This example clearly shows that a simplistic word count comparison between the ST and TT would not be sufficient to fully understand the phenomenon of TT expansion/contraction.

[5]
ST: 솟을 *대문을* 들어서면 안채와 사랑채, 가묘,
[*sosseul daemun-eul deureo-seomyeon anchae-wa sarangchae gamyo*]
'rising big gate-OBJ enter-stand-CDN women's sector-CNN men's sector' family graves'
동춘당 선생만을 위한 별묘 등이 배치되어 있다.
[*Dongchundang seonsaeng-man-eul uihan byeolmyo deung-i baechi-doe-eo issda*]
'proper name master-only-OBJ honoring Byeolmyo etc.- SBJ layout-PSS-CNN having been'
TT: If you pass through '*Soteuldaemun*' (the name of the main gate), you could meet four sub-buildings standing magnificently on the *Dongchungotaek* site: *Sarangchae* (used as a reception room for male guests), *Anchae* (for the female family members) and *Byeolmyo* (built to hold memorial services only for *Dongchundang*)[8].

Most intriguingly, example [6] presents a case of TT expansion driven by a Korean writing convention which allows for, or sometimes encourages, repetitive orthographic representations of the same information piece, a novel phenomenon which has not been reported thus far in TS, as far as the author is informed. In example [6], equipment component names of the ST (*The Cadex Analyzer* and *Cadex Batteryshop TM*) are rendered repetitively in the TT, first in Korean followed by their respective English correspondents put in brackets[9].

[6]
ST: The Cadex Analyzer family consists of the 2-bay Cadex 7200 analyzer and Cadex Batteryshop TM – a Windows-based software.
TT: Cadex 분석 장치 [Cadex Analyzer] 제품군은 윈도우 기반
[*Kadeks bunseok jangchi* [Cadex Analyzer] *jepumgun-eun uindou giban*]
'Cadex analyzing equipment [Cadex Analyzer] product group-SBJ Windows-based'
소프트웨어인 Cadex 배터리샵 TM [Cadex batteryshop TM] 으로 구성
[*soptweeo-in kadeks baeteori syap ti-em* [Cadex Batteryshop TM] *euro guseong*]
'software-being Cadex batteryshop TM [Cadex Batteryshop TM]-with being-composed'

This writing convention of Korean is motivated either by the desire to disambiguate homonyms of Korean or to give additional information when presenting concepts of foreign origin introduced in the ST.

From the above examples, it has been established that the following analyses are required: (1) exploration of TT contraction as well as TT expansion as phenomena which cannot be explained simplistically, and (2) implementation of a study designed in such a way as to guarantee investigation of all four aspects of the complex phenomenon in question.

2.2 Research Issues

2.2.1 Devising reliable measurement units

It should be clear by now that devising a set of reliable measurement units instead of a single one-size-fits-all analytical unit is crucial, in order to obtain an in-depth understanding of the complex translation behavior in question. In order to meet this challenge, the following four measurement units were devised:

(1) Word Count Rate (WCR), as the basic unit to track overall text length changes between the ST and TT
(2) Sentence-Initial Connectives Frequency Rate (CFR) in order to compare the respective frequencies of the cohesion device in the ST and TT to see any increase in the TT in the level of explicit text cohesion
(3) Parenthesis Frequency Rate (PFR), which tracks occurrences of supplementary explanatory information given in the TT in addition to what is already verbalized in the ST
(4) Bracket Frequency Rate (BFR), as a tool to locate repetitive orthographic representations of the same information.

WCR, the simplest measurement unit, is calculated by dividing the number of words (tokens) of the TT by the corresponding number in the ST. CFR is the ratio of the total occurrences of sentence-initial connectives of the TT as compared to those of the ST. In calculating PFR, all pieces of supplementary explanatory information that are not in ST but are present in TT were manually identified and put in parentheses in order to see the relative frequency of occurrences in the ST and the TT. As for BFR, all text segments with repetitive orthographic representations were manually identified as well, and marked with brackets (See 3.2 for text pre-processing).

2.2.2 Variable Candidates to be Analyzed

To repeat the purposes of this study for the sake of convenience, this study intends (1) to see whether either TT expansion or TT contraction is observable in translations involving English and Korean (2) to see whether the direction of language combination (Korean-into-English and English-into-Korean) has any influence upon the translation behavior in question, and (3) to identify possible factors intervening in TT expansion or contraction. In order to accomplish the third goal, a list of candidates of 'TT expansion factors' was constructed by means of a literature review, which is summarized as *Table 1.*

Table 1

Variable Candidates & Annotation Categories

	Annotation Categories	Sub-Categories						
Language	Lg. Combination Direction (Lcdr)	English-into-Korean (Ek)			Korean-into-English (Ke)			
Text	Text Function (Typ)	Informative (Inf)		Operative (Opr)		Expressive (Exp)		
	Subject Domain (Dmn)	Scnc/IT	Ecnm	P/Sct	Clt/Tr	Mdc/Hlth	Litr	
	Genre Membership (Gnr)	Advt	Nws	Rprt	Lgld	Stry	Thss	Mnl

Readership Relation[10]	Domain-specific Knowledge (Dмn-K)	Y-Y	Y-N	N-Y	N-N
	Culture-specific Knowledge (Cлт-K)	Y-Y	Y-N	N-Y	N-N
	Familiarity with Subject Matter of the Text (Fмl-Sм)	Y-Y	Y-N	N-Y	N-N

Abbreviations used in the Table

SCNC/IT: Science & IT	ECNM: Economy	P/SCT: Politics & society	CLT/TR: Culture & tourism
LITR: Literature	ADVT: Advertisement	NWS: News	RPRT: Report
MNL: Manual	LGLD: Legal document	STRY: Short story	THSS: Academic thesis
MDC/HLTH: Medicine & public health			

‘Readership relations’ is a concept newly coined by the author, referring to relative ‘presuppositions-in-situation’ of the ST readership and the TT readership. ‘Presuppositions-in-situation,’ as used here, are the aggregate of the cognitive knowledge one is assumed to have at the time of accessing the text as an object of communication. As Schmidt (1976, cited in Nord 1991: 41) notes, communication can only be successful if both the author and the reader implicitly assume the same presuppositions in sufficient quantity. Likewise, translation can be successful only if the TT reader has the same presuppositions-in-situation as the ST author assumes the ST reader possesses. During translation, a competent translator first gauges possible ‘overlaps and gaps’ in presuppositions-in-situation between the ST and TT readers, and makes necessary mediations accordingly. This in turn induces different degrees of mediation such as explicitation/implicitation, thus different levels of TT expansion/contraction.

To facilitate our understanding of the concept, let us assume a translation situation where the ST readers are experts in the subject domain of the ST but where the TT readers are only lay people. The translator judges that the target readership of the ST is a group of experts in a certain subject domain, who are assumed to have abundant knowledge of the subject matter. The basis of her judgment is the observation that the text is concise and to the point, without much explanatory information which might be viewed as redundant or even “condescending” (Leppihalme 1997: 110). On the other hand, she assumes that her TT readers are different. Lacking the same kind of presuppositions-in-situation as experts, the TT readers would find a TT less accessible if it lacked additional information to fill the gaps. In other words, gaps in presuppositions-in-situation between the ST and TT readers oblige the translator to actively intervene and make necessary mediation in rendering the TT.

To sum up, the term ‘readership relations’ signifies ‘a comparison of the ST and TT readers’ in terms of the presuppositions-in-situation assumed regarding what they will be able to invoke when attempting to understand the text. Readership relations can be compared in terms of (1) domain-specific knowledge (2) culture-specific knowledge, and (3) familiarity with subject matter of the text.[11] Hypothetically, the extent of TT expansion/contraction can vary, depending upon overlaps and gaps in presuppositions-in-situation possessed by ST and TT readers.

The assumption made is that TT expansion/contraction might be influenced by whether (1) both the ST and TT readers have expertise in the subject domain of the

text, (2) neither of them possesses this expertise, or (3) only one readership is knowledgeable of the subject domain. To make the comparison simple, let 'Y' signify those with presuppositions in the relevant subject domain and 'N' those without. Then a total of four combinations emerge: 'Y-Y' 'Y-N' 'N-Y' and 'N-N.'[12]

III. Construction of an Independent Parallel Corpus
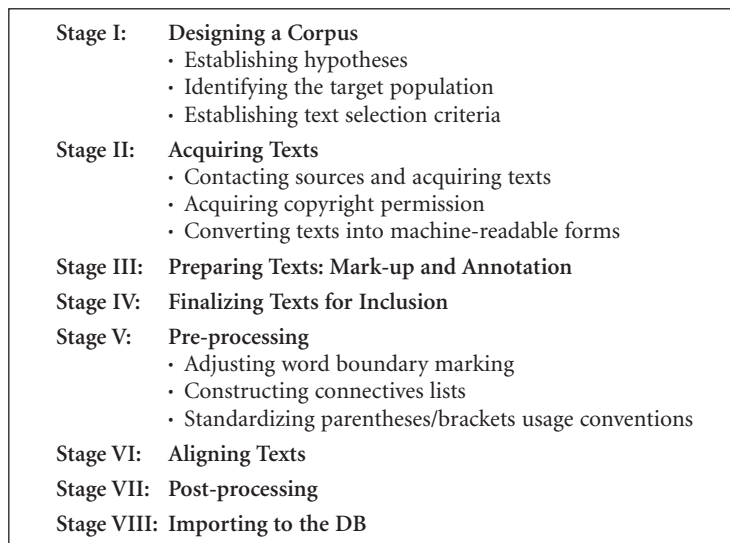
3.1 Designing a Corpus

For this study, a new corpus was compiled in eight stages (*Diagram 1*, See also Cheong 2004 for detailed descriptions of each stage). Deciding how large is large enough for a corpus under consideration to warrant valid generalizations lies at the center of corpus design. As noted by Bowker and Pearson (2002), however, there are no hard and fast rules that can be applied in determining the ideal size of a corpus. As a new corpus was compiled in order to supply empirical data for my research, and acutely aware of the time-, energy-, and thus money-consuming nature of the work involved, it was decided, in close consultation with statisticians, to maintain the corpus size at a manageable level as long as the total number of texts collected and the number of texts in different analysis unit groups could meet the statistical requirement for producing valid data and subsequent valid generalizations.

The target population of the corpus compiled is (1) naturally occurring, authentic translations realized as a result of actual translation needs (2) produced by professional translators (3) produced by Korean translators who are mature native speakers of Korean and non-native speakers of English (4) representing both literary and technical texts (5) synchronically produced (6) comprising both LGP texts (representative of everyday, non-specialized language) and LSP texts, and (7) bilingual (produced in one of the two languages, i.e., Korean and English) translations collected in a parallel corpus as a corpus of ST's in LA and their translations into LB.

Diagram 1

Corpus Compilation Flowchart

Stage I:	**Designing a Corpus**	
	· Establishing hypotheses	
	· Identifying the target population	
	· Establishing text selection criteria	
Stage II:	**Acquiring Texts**	
	· Contacting sources and acquiring texts	
	· Acquiring copyright permission	
	· Converting texts into machine-readable forms	
Stage III:	**Preparing Texts: Mark-up and Annotation**	
Stage IV:	**Finalizing Texts for Inclusion**	
Stage V:	**Pre-processing**	
	· Adjusting word boundary marking	
	· Constructing connectives lists	
	· Standardizing parentheses/brackets usage conventions	
Stage VI:	**Aligning Texts**	
Stage VII:	**Post-processing**	
Stage VIII:	**Importing to the DB**	

The sampling procedure was guided by the 'motivated choice' principle in order to represent the target population delimited above. The attributes of that population were adopted as text selection criteria. A total of 158 text pairs were ultimately included. All texts collected were full or complete texts only, rather than "excerpts" (Laviosa 1997; Olohan 2002), which might "accidentally eliminate a part of the text that could be very interesting" (Bowker & Pearson 2002: 49). Accordingly, short stories were selected to represent literary texts. During the selection process, efforts were made to maximize the diversity of the texts included.

3.2 Text Preparation, Pre-Processing, Text Alignment, and Post-Processing

A text collection, when completed, underwent mark-up (a process to give information to texts in order to explicitly name the different sections of a document) and annotation (a process which makes the linguistic features of a text explicit by adding information about texts (Bowker & Pearson 2002)). The annotation work[13] for this study was done by assigning membership in one of the seven categories to each text, followed by assignment to a sub-category group (See *Table 1*.)

Also efforts were made to enable the Database to automatically scan the corpus and identify corresponding occurrences of each measurement unit: WCR, PFR, BFR, and CFR. For CFR, respective lists of Korean and English connectives were constructed since the data in the corpus is not P-O-S (part of speech) tagged[14]. The current level of computation technology does not allow for automatic scanning and location of information given in TT in addition to what is in the corresponding ST, nor can incidences of repetitive orthographic representations be automatically noted. Because of this limitation, brackets and parentheses to respectively mark repetitive orthographic representations and extra information supplied in TT's were inserted manually. This permitted the calculation of BFR and PFR.

After the selection of texts and the subsequent pre-processing, collected texts were aligned in such a way as to make it easier to look at translations with relation to their respective STs and to allow for targeted item(s) to be called up on screen in order to check them against the corresponding ST or TT item (Munday 1998).

The text alignment for this study, with sentences as the intended level of alignment, was preliminarily done using the WinAlign program, a UNICODE-based component of the TRADOS program, which supports Asian languages including Korean[15]. The final alignment was completed after the end-result of the preliminary alignment, a set of intercalated texts (created by inserting the corresponding sentence of the TT after each sentence of the ST), was imported to the Database built for this study. A total of three rounds of both automatic and manual checks were done before the completion of the alignment: manually by the WinAlign programmer and the author, one automatic round by the DB by comparing the numbers of the ST and TT sentences, and a final manual correction. *Diagram 2* shows the finalized aligned text sample.

Diagram 2
Aligned Text Sample
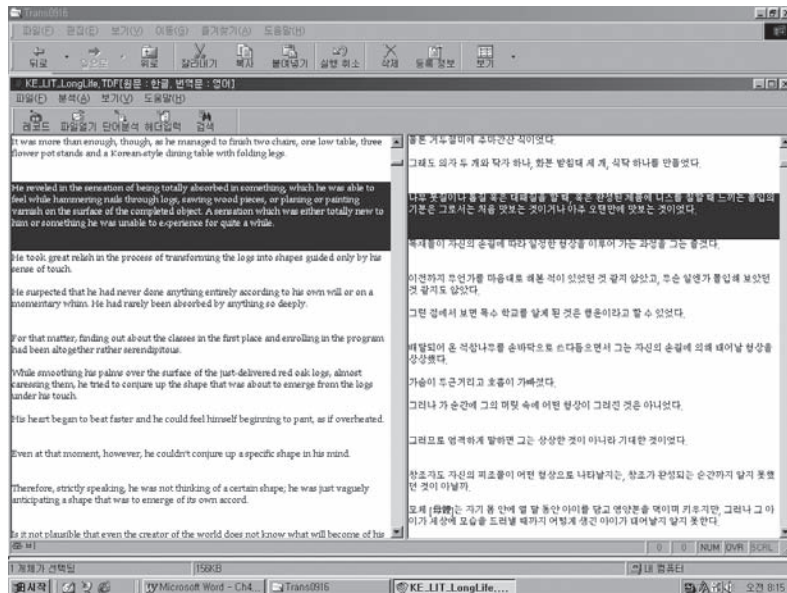
IV. Statistical Analyses

After the parallel corpus was ready for statistical treatment, texts in the database were automatically scanned first in order to identify the number of words and sentences, the occurrences of sentence-initial connectives, and the occurrences of information enclosed in either parentheses or brackets in accordance with the standardized parenthesis/bracket marking of the ST and TT in pairs. The result was a calculation of the corresponding TT expansion rates for each of the text pairs for all four measurement units: WCR, CFR, PFR, and BFR. Then TT expansion/contraction rate means were calculated for texts belonging to the same annotation groups (e.g., texts with the INF text function type and the legal document genre) (*Diagram* 3).

With the target text expansion rates[16] (the degree of TT expansion/contraction in comparison to the ST) of individual texts and the expansion rate group means of texts belonging to the same group of an annotation category, one-way analyses of variance (one-way ANOVA) tests were conducted to identify significant variables as TT expansion factors. For this, *p-value* was used. *P-value* refers to the minimum level of significance probability, based on which a null hypothesis can be rejected. If the *p-value* of a variable candidate is equal to, or lower than, the predetermined significance level, the corresponding variable candidate can be judged to be a significant variable of the phenomenon under analysis. The 'confidence level' chosen for this study, which gives an estimated range of values that is likely to include an unknown population parameter, is 0.05, or 5%.

The analysis accomplished the following: (1) determination of whether TT contraction as well as TT expansion actually occurred in the texts in different annotation

categories, (2) identification of significant variables of the phenomenon, or expansion factors, (3) identification of actual variance in TT expansion rates in different groups of the annotation category, and (4) identification of text groups with the largest and smallest TT expansion rates.

DIAGRAM 3

Text Analysis Results

총 레코드 갯수 : 158

| File Name	FNC	SL	TL	WCR	eWC	kWC	C
KE_PSJ_Yonhap_0801.TDF	INF	K	E	1.498	406	271	1.0
KE_PSJ_SegyeInterview.TDF	INF	K	E	1.365	213	156	0.0
KE_PSJ_Munwha_0313.TDF	INF	K	E	1.305	338	259	2.3
KE_PSJ_KimdaejungInterview.TDF	INF	K	E	1.484	711	479	1.4
KE_PSJ_HankookEcoDaily_1221.TDF	INF	K	E	1.341	283	211	1.0
KE_PSJ_Hankook.TDF	INF	K	E	1.647	140	85	0.0
KE_PSJ_Hangyeoreh.TDF	INF	K	E	1.398	158	113	3.0
KE_PSJ_DongAllbo_0417.TDF	INF	K	E	1.405	680	484	2.5
KE_LIT_LongLife.TDF	EXP	K	E	1.873	13095	6992	2.0
KE_LIT_BuckWheat.TDF	EXP	K	E	1.695	3550	2095	5.0
KE_LIT_BareHills.TDF	EXP	K	E	1.624	2069	1274	1.9
KE_LIT_Adada.TDF	EXP	K	E	1.549	3410	2202	2.3
KE_Kyemyong_Toegye.TDF	INF	K	E	1.450	319	220	0.5
KE_Kwangju2.TDF	INF	K	E	1.675	7367	4399	1.4
KE_Kwangju1.TDF	INF	K	E	1.637	5195	3174	1.0
KE_CHJ_KoreanClothes.TDF	INF	K	E	2.082	1757	844	1.2
KE_CHJ_EcoTour.TDF	OPR	K	E	1.707	2445	1432	1.2
KE_021112_ASE_Dz20.TDF	INF	K	E	1.591	210	132	0.0
KE_021111_맨앤툴_Cb8.TDF	OPR	K	E	1.346	2088	1551	2.2

Save   선택

V. Findings and Discussions

5.1 TT Contraction

To sum up the findings, the first thing to report is that a distinctive tendency toward shorter TTs, or TT contraction, has been identified in tandem with the often-noted TT expansion. This finding is generalized across the data, and is not restricted to results based on only one type of measurement.

Most notably, the translating language combination direction (LCDR), i.e., English-into-Korean (EK) and Korean-into-English (KE), was found to have a significant bearing upon the extent of TT expansion/contraction. The most salient finding is that the combination of English-into-Korean leads to TT contraction (TT expansion rate of 0.863) and Korean-into-English to TT expansion (1.490) when measured in WCR (*Figure 1*).

TT Expansion Rate Means, by LCDR (All Data, in WCR)
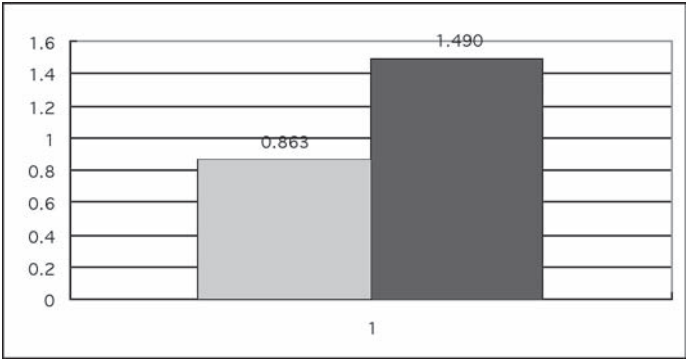
Until now, in the TS literature, TT expansion has been considered the most likely potential candidate to be included in any inventory of TU's. Finding such strong support in the form of almost unanimous reports in the research literature, TT expansion has begun to be widely accepted as a universal translation phenomenon, regardless of the differences in the translating language combinations. Nord (1996) and Vanderauwera (1985) have noted, however, that certain ST text segments can be omitted or deleted in the TT in cases where they represent repetitive or redundant information, and that this in turn may result in local contraction in the text. However, few – if any – reports have been made of the contraction of the TT as a whole. This study offers counter-evidence arguing against the prevailing belief in the universality of explicitation and the resultant TT expansion.

5.2 TT Expansion Factors Verified, by Measurement Unit

With WCR, CFR, BFR, and PFR as four measurement units with which to track quantitative changes between the ST and TT, the actual list of significant variables, or 'TT expansion factors,' in accounting for of the translation phenomenon in question are verified. *Table 2* summarizes the respective *p-values* of all variable candidates proposed in this study, from which we can identify significant variables of the TT expansion/contraction. The significant variables are marked with an asterisk next to corresponding *p-values*.

TABLE 2

P-Values of Candidate Variables, by Measurement Units
(Significance Level = 0.05)

Measurement Units	Candidate Variables	p-Value		
		All Data	KE	EK
WCR	1. Lg. Combination Direction (LCDR)	0.000*		
	2. Text Type (TYP)	0.092	0.075	0.000*
	3. Text Subject Domain (DMN)	0.000*	0.000*	0.030*
	4. Text Genre (GNR)	0.039*	0.363	0.000*

	5. Presupposed Domain-specific Knowledge (DMN-K)	0.002*	0.000*	0.004*
	6. Presupposed Culture-specific Knowledge (CLT-K)	0.000*	0.002*	0.576
	7. Presupposed Familiarity with Subject Matter (FML-SM)	0.000*	0.005*	0.224
CFR	1. Lg. Combination Direction (LCDR)	0.000*		
	2. Text Type (TYP)	0.997	0.674	0.344
	3. Text Subject Domain (DMN)	0.499	0.882	0.385
	4. Text Genre (GNR)	0.224	0.575	0.000*
	5. Presupposed Domain-specific Knowledge (DMN-K)	0.878	0.867	0.304
	6. Presupposed Culture-specific Knowledge (CLT-K)	0.329	0.612	0.710
	7. Presupposed Familiarity w/Subject Matter (FML-SM)	0.373	0.797	0.269
PFR	1. Lg. Combination Direction (LCDR)	0.173		
	2. Text Type (TYP)	0.017*	0.043*	0.004*
	3. Text Subject Domain (DMN)	0.083	0.008*	0.044*
	4. Text Genre (GNR)	0.098	0.093	0.026*
	5. Presupposed Domain-specific Knowledge (DMN-K)	0.430	0.725	0.536
	6. Presupposed Culture-specific Knowledge (CLT-K)	0.172	0.092	0.117
	7. Presupposed Familiarity w/Subject Matter (FML-SM)	0.356	0.356	0.472
BFR	1. Lg. Combination Direction (LCDR)	0.000*		
	2. Text Type (TYP)	0.495	0.100	0.453
	3. Text Subject Domain (DMN)	0.112	0.001*	0.518
	4. Text Genre (GNR)	0.614	0.345	0.494
	5. Presupposed Domain-specific Knowledge (DMN-K)	0.036*	0.070	0.137
	6. Presupposed Culture-specific Knowledge (CLT-K)	0.317	0.003*	0.861
	7. Presupposed Familiarity w/Subject Matter (FML-SM)	0.121	0.079	0.625

5.3 Language Combination Direction (LCDR)

What can be noted from the list above is that the variables which proved to be significant in accounting for TT expansion/contraction are different, depending upon the direction of the working language combination: Korean-into-English and English-into-Korean. It was found that depending upon different LCDR subgroups, different factors proved to be significant to differing degrees in accounting for TT expansion/contraction. To be specific, when measured in WCR, LCDR, text subject domain (DMN), presupposed culture-specific knowledge relations (CLT-K), presupposed familiarity with subject matter (FML-SM) (all with *p-values* of 0.000), and presupposed domain-specific knowledge relations (DMN-K) (0.039) were significant variables of the TT expansion/contraction between ST and TT involving Korean and English as working languages in general.

However, when LCDR is factored in, a different set of factors emerged to be significant in each language combination direction (*Table 3*). Specifically, text subject

domain (DMN) and presupposed subject domain-specific knowledge relations (DMN-K) (*p-values* of 0.000), presupposed culture-specific knowledge relations (CLT-K) (0.002), and presupposed familiarity with subject matter relations (FML-SM) (0.005) were found significant in the KE Data group. In the EK Data group, on the other hand, text function (TYP) and genre (GNR) (*p-values* of 0.000), in addition to presupposed domain-specific knowledge (DMN-K) (0.004) and text subject domain (0.030) were found to be significant. Text function (TYP) and text genre (GNR) were not found to be significant in the KE Data group (*p-values* of 0.075 and 0.363 respectively), contrary to the findings for the EK Data group (both 0.000). These results are presented in *Table 3*.

TABLE 3

TT Expansion Factors, by Language Combination Direction

Lg. Combination Direction	Significant TT Expansion Factors (in WCR)
Korean-into-English	DMN & DMN-K (both *p value*=0.000) > CLT-K (0.002) > FML-SM (0.005)[17]
English-into-Korean	TYP & GNR (both *p value*=0.000) > DMN-K (0.004) > DMN (0.030)

What these results tell us is that even in translations with the same translating language combination, depending upon the direction of that combination, different factors intervene in the TT expansion/contraction process, exerting different degrees of influence.

To further verify whether language combination direction is a significant expansion factor, another test, the *t*-test, was conducted. Levene's test was used to verify the equality of variances between the two groups (EK and KE), using "H0 = There is equal variance in both EK and KE groups." The test result was F = 2.946, the significance level of which is .088. Since the significance level chosen was .05, the null hypothesis cannot be rejected. Therefore, equal variance can be assumed of the groups. In the t-test, with equal variance assumed, "*H0 = There is no WCR differences between EK and KE groups*." The test result was T = -2.223, with 146 degrees of freedom,[18] the significance level of which is .028. Since the significance level calculated is smaller than the significance level selected, the null hypothesis can be rejected (*Table 4*).

TABLE 4

T-test between EK and KE group (in WCR)
(Significance level of .05)

Measurement unit	H0	Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t-value	Df	Sig. (2-tailed)
WCR	Equal variances assumed	2.946	.088	-2.223	146	.028

Therefore, it can be concluded that there are differences in WCR between EK and KE groups.

The fact that language combination direction (LCDR) is an actual expansion factor when measured in WCR, CFR, and BFR (all with *p*-values of 0.000) raises a warning signal to the TS discourse, which has placed more focus on 'intercultural differences' or 'differences in the shared cognitive environment.' Little attention has thus far been placed on the characteristics of the languages *per se* involved in the translation and the derivative characteristics that emerge from the combination direction of the working languages. In this regard, the finding of this research is meaningful.

This should not be construed to mean, however, that languages alone dictate the translator intervention in rendering the TT. Language combination directions interact with other TT expansion factors, producing variations in translation behavior of texts with relation to TT expansion/contraction (See Cheong 2004 for details.).

5.4 Four Measurement Units

The four measurement units devised for this study proved to be instrumental in tracking TT expansion/contraction as a complex phenomenon.

First, Word Count Rate (WCR) proved to be useful in tracking the word count changes between the ST and TT based on the comparison of the number of lexical words which respectively constitute the ST's and the TT's. However, the weakness of this measurement unit was acknowledged from the start of the research. Put differently, this method alone cannot provide a detailed picture of the TT expansion/contraction phenomenon since an expansion in one text segment can be offset by a contraction in another segment. It is also possible that several small expansions can be offset by a single, large contraction, thus obstructing an accurate analysis of the phenomenon.

Second, Sentence-Initial Connectives Frequency Rate (CFR) allowed for a more focused look at changes in the occurrence frequencies of the cohesive link device. With this data, it was possible to understand one of the main motivations behind the TT expansion/contraction phenomenon, responding to different conventions of the two language communities involved regarding the use of the cohesion device. Traditionally, it has been assumed in the literature that the use of connectives in general and of sentence-initial connectives in particular increases due to the perceived need for explicitation, or a tendency to make implicit ST information explicit in the TT (Vanderauwera 1985, Baker 1996, Gutt 1992, Nord 1996, *inter alia*). Against this backdrop, it is noteworthy that the CFR for the English-into-Korean Data group represents a contraction, albeit a very slight one, standing at 0.987.

A plausible explanation for the apparent lesser reliance upon explicit sentence-initial connectives can be found in one of the characteristics of the Korean language. Korean is a language with a peculiar class of auxiliary suffixes, or bound morphemes, which are attached mostly to nouns, and which add various meaning, among which are case marking and clause connections. Because of this, linking adjacent clauses or sentences in Korean in order to enhance cohesion of the text does not rely solely on independent words, or *jeopsok-busa* 'connective adverbs,' but also on these bound morphemes. The relatively frequent use of the lexical class of bound morphemes results in less use of explicit sentence-initial connectives in Korean than in English. As a result, the CFR-based EK Data group mean registered a contraction, albeit

minimal. In this regard, CFR points to the differences in the translation behavior between the English and Korean language, which stem from different characteristics of the languages.

With PFR, it was possible to identify text segments where extra explanatory information is intentionally supplied in the TT's, as well as cases where explicitly verbalized information in the ST is omitted in the corresponding TT segment. When PFR was used as a measurement unit, no contraction was observed in the group means of All Data (1.183), EK (1.347), and KE (1.018) groups. The fact that only expansion is observed in all three data groups is consistent with the common sense idea that translation generally involves the supply of explanatory information in text segments that are new or unfamiliar, rather than the deletion of information that is already old. The assumption underlying this phenomenon is that the presuppositions-in-situation of the TT readership is supposed to be more limited than that of the ST readership.

Last but not least, the Bracket Frequency Ratio (BFR) proved to be a reliable device with which to look at another feature of TT expansion/contraction whereby already explicitly verbalized information is repetitively given in a different ortho-graphic representation. Translation behaviors as such involve the representation of certain concepts or proper names in two different languages, e.g., Korean and Chinese, and Korean and English. This is not solely for informativity equivalence-related reasons. Rather it is related to the convention – not only in translation but also in writing in general – that acronyms are normally spelled out in their first appearance in a text, or that proper names of firms, organizations, products, or individuals of foreign origin tend to be followed by their original representation. What norm translators actually abide by in relation to this issue differs from one language community to another. The possible renderings of "IPO" into Korean, for instance, are: (1) "*gi-eop-gong-gae*" 'public offering of a company,' (2) "IPO" (3) "*gi-eop-gong-gae* [IPO]" (4) "*gi-eop-gong-gae* [Initial Public Offering]," and (5) "*gi-eop-gong-gae* [Initial Public Offering], [IPO]." Depending upon the choice made by the translator, the frequency of the occurrence of information placed in brackets and thus of the overall length of the TT will vary.

Other than conformity to convention, the translator's choice can be influenced by her assumptions about the shared cognitive environment (Sperber & Wilson 1986; Gutt 1992) of the TT readership. In simpler terms, the translator, driven by a perceived need to enhance accessibility in rendering the TT, chooses an option which she deems warrants the same degree of informativity as in the ST. This choice is in turn closely related to her assumptions about the size of the presupposition-in-situation of the TT readership. If the translator assumes that her TT readership is one with expertise in the subject domain of the text being translated, and thus familiar with the acronyms in the field, she will not resort to giving repetitive information in different orthographic renderings. However, if the TT reader is assumed to be a lay person, the translator would more often exploit the tactic of encoding the same information repetitively in different forms. When approaching the TT expansion/contraction phenomenon in terms of BFR, it was found that BFR is higher in texts translated for a TT readership of laypersons, compared to texts rendered for TT readers with expertise in the subject domain of the text. This leads us to conclude that BFR is valuable in allowing us to understand another aspect of the phenomenon in question.

To summarize, the four measurement units proposed in this study proved effective and significant in accounting for explicitation/implicitation and the subsequent TT expansion/contraction. Word Count Rate was especially significant in that the largest number of factors was found significant when measured with this particular yardstick. The significance of this unit was already pointed out by Johansson (1995, cited in Baker 1996), who reported an increase of approximately 10% in the number of words in English translations vs. Norwegian originals. However, its validity was even more salient in this study, for TT contraction was observed in English-into-Korean translations[19] when measured with that tool.

On the other hand, other three measurement units – the Sentence-initial Connectives Frequency Rate (CFR), the Parenthesis Frequency Rate (PFR), and the Bracket Frequency Rate (BFR) – also proved to be effective and valid in explaining the phenomenon under investigation, producing a list of significant TT expansion factors verified. However, they are also valuable in accounting for what possibly motivates the explicitation/implicitation and TT expansion/contraction other than language-related variables. Put differently, when measured in WCR, TT expansion/ contraction occurs 'as a function of the variables' that have proved significant in the study, including language combination direction (LCDR). Secondly, TT expansion/contraction phenomena, measured in CFR, are found to be related to differences between the two translating languages in 'the convention governing the reliance upon explicit use of sentence-initial connectives.' Thirdly, the phenomenon tracked with PFR is attributable to 'a need to enhance accessibility of the TT.' Lastly, the differences in the conventions pertaining to the repetitive orthographic representation of the TL expression in tandem with the SL expression can be understood in relation to 'differences in the text production conventions' between the English and the Korean language communities. It seems that Korean allows for, or even encourages in some cases, the active use of foreign language representations such as Chinese or English, in facilitating disambiguation or providing additional information when deemed necessary.

5.5 Readership Relations as Novel Variables

One of the important motivations of this study has been to introduce a set of 'novel variable candidates' as potential variables of the TT expansion/contraction phenomenon, and to test their validity as significant expansion factors in accounting for the phenomenon. As opposed to 'traditional variables,' a term used to refer to variables on which scholarly attention has focused and which include such factors as text type, subject domain, or text genre, the proposed 'novel variables' are presuppositions-in-situation which the ST and the TT reader are assumed to possess in terms of (1) subject domain-specific knowledge relations (DMN-K), (2) culture-specific knowledge relations (CLT-K), and (3) familiarity with subject matter (FML-SM).

The results of the one-way ANOVA tests were that all three proposed variables were found to be significant in some data groups, as summarized in *Table 5*. From these findings, it can be safely proposed that these novel variables should be introduced and exploited in research on translation behavior of texts.

Table 5
Significance of Novel Variables, by Data Group

Measurement unit	Novel variables	Data groups in which 'novel variables' proved significant		
WCR	DMN-K	All (*p-value* of 0.002)	KE (0.000)	EK (0.004)
	CLT-K	All (0.000)	KE (0.002)	-----
	FML-SM	All (0.000)	KE (0.005)	-----
BFR	DMN-K	All (0.036)	-----	-----
	CLT-K	-----	KE (0.003)	-----
	FML-SM	-----	-----	-----

The initial intention of this study was to subsume these three variables under a more generic, 'super' variable. However, the data collected were not evenly distributed among the groupings. In addition, several groups had texts in insufficient quantity to guarantee reliable generalization. What this suggests is that now that the tentative validity of the three newly introduced variables have been verified by the current study, further research efforts should be made to obtain a statistically sufficient number of texts in each group, thus allowing for the three-way cross-analysis of the measure of association among these factors, using λ-values.

Table 6
Measure of Association among Novel Variables[20]

Variable	Groups	Highest association	2nd-highest association	3rd-highest association
DMN-K	All-Data	FML-SM 0.319 (0.02)	GNR 0.285 (0.000)	CLT-K 0.252 (0.022)
	EK	GNR 0.289 (0.005)	FML-SM 0.267 (0.028)	-----
	KE	FML-SM 0.370 (0.001)	DMN 0.354 (0.000)	CLT-K 0.347 (0.009)
CLT-K	All-Data	DMN 0.281 (0.000)	DMN-K 0.252 (0.022)	FML-SM 0.231 (0.002)
	EK	-----	-----	-----
	KE	FML-SM 0.489 (0.000)	DMN 0.371 (0.000)	DMN-K 0.347 (0.009)
FML-SM	All-Data	DMN-K 0.319 (0.002)	CLT-K 0.231 (0.004)	-----
	EK	DMN-K (0.267) 0.028	-----	-----
	KE	CLT-K 0.488 (0.000)	DMN-K 0.370 (0.001)	GNR 0.252 (0.005)

As is summarized in *Table 6*, the three novel factors, i.e., presupposed subject domain-specific knowledge relation (DMN-K), presupposed culture-specific knowledge relation (CLT-K), and presupposed familiarity with subject matter relation (FML-SM), not only significantly influence TT expansion/contraction but also exert influence upon one another.

The significance of the three variables is even more striking when it comes to the KE translations. First, the three variables are found to have a significant degree of

association among themselves. Second, the strength of association among them is quite robust, ranging from 0.347 to 0.489. This finding points to the need to seriously consider these variables as expansion factors, particularly in the case of KE translations.

Also the finding can be regarded as supporting the need to incorporate the extent to which the TT readership is exposed to the SL culture. In the case of the KE Data group, CLT-K and presupposed familiarity with subject matter (FML-SM) show 'relatively strong' association, with a -value of 0.489, a much stronger association than in the EK data group. This finding is consistent with the generally held assumption that English-medium culture values are better known in Korea than vice versa. In other words, even though the same languages are used in a specific translation, the extent that one culture is exposed to the other and the perceived familiarity the TT readers have vis-à-vis the SL culture should be taken into account when explaining TT expansion/contraction.

Concluding the findings pertaining to the three novel variables, it can be said that these three variables proved to be significant in accounting for the phenomenon of text expansion/contraction.

VI. Conclusion

6.1 Summary

This study explored TT expansion and contraction as realizations of explicitation and implicitation, respectively. The significant findings of this study are that (1) TT contraction as well as much-noted TT expansion were observed in translations involving Korean and English, and (2) the actual tendency toward TT expansion and TT contraction was influenced by many 'TT expansion factors,' including the direction of translating language combination. With this, it can be safely concluded that (1) explicitation, and TT expansion as its text-level realization, is not necessarily predominant or even present in all translations, and (2) the direction of translating language combination leads to a different TT expansion/contraction behavior of texts even in cases where the identical language pair is involved.

The above findings are empirically grounded and result from an in-depth analysis of a specially-constituted parallel corpus of translations between Korean and English consisting of 517,609 token words (299,630 English and 218,979 Korean) in 322,945 sentences in 316 authentic texts. Four different measurement units were devised and their validity was verified in this study: (1) Word Count Rate (WCR), as the basic unit to track overall text length changes between the ST and the TT, (2) Sentence-Initial Connectives Frequency Rate (CFR) in order to compare the respective frequencies of the cohesion device in the ST and TT to measure any increase in the TT with respect to the level of explicit text cohesion, (3) Parenthesis Frequency Rate (PFR), which tracks occurrences of supplementary explanatory information given in the TT in addition to what is already verbalized in the ST, and (4) Bracket Frequency Rate (BFR), as a tool to locate repetitive orthographic representations of the same information. The motivation for employing diverse measurement units was to capture diverse aspects of the phenomenon in question, which I believe to be more complex and comprising more factors than anticipated in the literature.

A set of newly devised variables – presupposed domain-specific knowledge (DMN-K), presupposed culture-specific knowledge (CLT-K), and presupposed familiarity with

subject matter of the text (FML-SM) – proved to be significant in accounting for TT expansion/contraction. The underlying assumption of the validity of the 'novel variables' is that the extent of explicit verbalization of texts hinges upon the amount of the presuppositions-in-situation the author assumes of his readership, and that differences between the author of the ST and the translator in their assumptions regarding their respective readership's presuppositions-in-situation leads to different levels of explicit verbalization of the text, leading to different extents of TT expansion/contraction.

6.2 Significance of the Present Study and Future Study Needs

Ever since the incorporation of 'Cultural Turn' into TS, most of the research attention has focused on exploring culture-related issues. This translates into less emphasis on language-related considerations. However, it was shown in this study that the translating languages and the direction of language combinations, even in cases where the same languages are involved, exert influences upon the translation behavior of texts under investigation. This would point to a need to systematically assess the significance/function of specific working languages involved in future research efforts.

The findings of the present study have a direct bearing upon what to teach and how to teach it in translation classrooms as well. They support the need to raise the awareness of TT expansion/contraction in the minds of students of translation as a consequence of translator mediation. Students should be explicitly taught to identify relevant expansion factors, depending upon the specific translation task, and translating language combination direction. Being familiar with various mediation devices, students should then be able to prioritize different translation tactics to suit the TT readership they assume for their task at hand. They should also learn what to consider in making decisions on the necessary degree of explicit verbalization of the TT. A TT which is too 'friendly,' information-wise, or a TT with too much 'kind' extra information supplied in addition to what is already in the ST, may enhance the accessibility of the target readership to the TT, but at the same time it runs the risk of looking 'condescending' to the target readership . In other words, the extent of TT expansion/contraction is related to the perceived quality of the TT in the minds of its end-users.

This leads to another issue to be taken up in future research. The present study did not address the appropriate level of explicitation/implicitation and the qualitative contents of such mediation. The verification that both TT expansion and TT contraction were quantitatively observed was significant. What is needed in the future is to determine what constitutes the criterion for the translator to use in making their decision on how to do the actual explicitation/implicitation mediation in the most effective way: rendering all that is needed to be rendered in the most economic way without either excessiveness or redundancy. This can be done using qualitative approaches.

All in all, future studies could usefully be devoted to the verification of the validity of the 'novel variables' against different sets of data, with the goal of proposing a 'super variable,' a variable incorporating all three variables and thus allowing for an easier, more comprehensive understanding of the TT expansion/contraction phenomenon motivated by the differences in the presuppositions-in-situation of the ST and TT readership.

NOTES

*   This work was supported by Hankuk University of Foreign Studies Research Fund of 2005.

1.   The sole comparable corpus in Korea is being built as part of the first stage of the Korean govern-
ment-funded 21st Century *Sejong* Project, which the Korean government launched in 1998 as a 10-
year long-term national project. However, only 20, albeit long, texts have been collected in the
parallel corpus as of the end of 2003.

2.   Romanization follows the prevailing system in Korea, which was officially endorsed as of July 2000
by the Korean Ministry of Culture and Tourism, and provides primarily a phonetic representation.

3.   Abbreviations used here are (in alphabetical order):

ADJ: Adjectival suffix	OBJ: Object case marker
ADV: Adverbial suffix	PRT: Present tense
CNN: Connectives	PSS: Passive voice
END: Verbal ending	PST: Past tense
HNR: Honorific marker	SBJ: Subject case marker

4.   Appropriateness or effectiveness assessment is not made here with relation to the actual contents of
the information supplied in addition to what is already in the ST.

5.   Efforts have been made to maintain consistency in the terminological usage: explicitation and
implicitation; expansion and contraction; and, addition/insertion and omission/deletion. The basic
assumption of such a decision is the belief that these concept pairs belong to three different layers
(See Cheong 2004 for detailed discussion.). Specifically, explicitation and implicitation lie at the
deepest layer, in the black box of the translator, motivating her to make necessary adjustments to
render the ST segment in such a way as to be accessible to the TT readership. Addition/insertion and
omission/deletion are at the intermediate layer, serving as the specific vehicles with which to make
that adjustment to the text. Finally, TT expansion and contraction lie at the top layer as the text-
surface realization of explicitation and implicitation.

6.   Source: Asia Ombudsman Association. Translation commissioner: the Korean Ministry of Culture
and Tourism.

7.   Source: *Rodong* Daily, North Korea's state-owned newspaper. Translation commissioner: the Korean
Ministry of Foreign Affairs and Trade.

8.   Source: Introduction to Korean Cultural Heritage. Translation commissioner: *Woorim* Corporation.

9.   In the original, this is given in parentheses. However, a systematic conversion was made to using
brackets since parentheses were already being used to mark information given in addition to what
is already in the ST (See 3.2. for text pre-processing).

10. As for annotating texts for the three categories under the 'readership relations' concept, three doctoral
students including the author with five or more years of practical experience, all familiar with the
relative distance between ST and TT readership in terms of presuppositions-in-situation, gave their
respective judgment independently. The final value of each text with relation to these categories was
established based on the majority principle.

11. This is a development of Nord (1991b), who listed temporal, spatial, cultural, and domain-specific
knowledge as components of one's presuppositions. Also see Cheong (2004) for the rationale for the
choices of these particular three as analysis units.

12. The annotation notations for this category are as follows:

'Y-Y'	The readership relation where both the ST and TT readership are assumed to have subject domain-specific knowledge pertaining to the subject matter
'Y-N'	The readership relation where the ST readership is assumed to have subject domain-specific knowledge which the TT readership is assumed not to have
'N-Y'	The readership relation where the ST readership is assumed not to have subject domain-specific knowledge which the TT readership is assumed to have
'N-N'	The readership relation where neither the ST nor the TT readerships are assumed to have subject domain-specific knowledge

13. The annotation process was guided by two principles: apply the 'prototype-based categorization
system' during the annotation of texts, and maximize the inter-subjective agreement on assigning
text membership status in sub-category groups to texts.

14. It was possible since connectives in a language constitute 'a closed category.' The lists of sentence-
initial connectives of Korean and English were constructed based on Kang et al. (2000), Sohn and

Haig (1997), and Nam and Ko ([1985] 1993) for Korean, and Frodesen and Eyring (1993) for English.

15. Other software programs, including Word-Smith and KWIC, developed primarily for automatic alignment and concordancing and which are available for free Internet download, could not be used due to their failure to support Korean.

16. The term denotes TT contraction as well as TT expansion. TT is considered to undergo TT expansion if the value is greater than 1, and TT contraction if the corresponding value is smaller than 1. Its value is calculated as [the length of TT measured in measurement unit x] ÷ [the length of ST measured in measurement unit x].

17. Here '>' symbolizes a stronger probability only; the strength of the two figures before and after the symbol cannot be directly compared.

18. Degrees of freedom (*df*) denotes the measure of variability, or the number of options available within a variable. In statistics of the N cells of a table of probabilities only N-1 can be arbitrarily filled: the degrees or freedom is N-1 (Easton & McColl, n.d.).

19. Johansson reported a 'slight' average increase in the number of words in the Norwegian translations compared to their English ST's.

20. The figures in each cell denote the value found at the intersection of the vertical and horizontal axes and the two corresponding variables' λ-value. The figures in parentheses are corresponding significance levels. λ with a value of '0' signifies the statistical independence of one independent variable from the other, and λ with a value of '1' denotes the complete dependence of the two independent variables involved.

REFERENCES

Baker, M. (1993): "Corpus Linguistics and Translation Studies: Implications and Applications," *Text and Technology: In honour of John Sinclair* (M. Baker *et al.*, eds.), Amsterdam and Philadelphia, John Benjamins, p. 223-250.

—— (1996): *Corpus-based Translation Studies*: *The Challenges that Lie Ahead*, H. Somer.

Blum-Kulka, S. (1986): "Shifts of Cohesion and Coherence in Translation," *Interlingual and Intercultural Communication* (J. House and S. Blum-Kulka, eds.), Tübingen, Gunter Narr, p. 17-35.

Bowker, L. and J. Pearson (2002): *Working with Specialized Language: A Practical Guide to Using Corpora,* London and New York, Routledge.

Cheong, H-J. (2004): *Translated Text Expansion and Contraction Phenomena: A Corpus-based Study of Quantitative Target Text Changes as Reflective of Translator Mediation,* Unpublished Doctoral Dissertation, Hankuk University of Foreign Studies, Korea.

Fillmore, C. (1982): "Frame semantics," *Linguistics in the Morning Calm* (Linguistic Society of Korea, ed.), Seoul, Hanshin, p. 111-138.

—— (1985): "Frames and the Semantics of Understanding," *Quaderni di Semantica* 6-2, p. 222-253.

—— (1992): "'Corpus Linguistics' vs. 'Computer-aided Armchair Linguistics,'" *Directions in Corpus Linguistics,* Proceedings from 1992 Nobel symposium on corpus linguistics, Berlin, Mouton de Gruyter, p. 35-60.

Frodesen, J. and J. Eyring (2000): *Grammar Dimensions: Form, meaning and Use,* Boston, Heinle & Heinle.

Gutt, E-A. (1990): "A Theoretical Account of Translation-without a Translation Theory," *Target* 2-2, p. 135-164.

—— (1992): *Relevance Theory*: *A Guide to Successful Communication in Translation,* New York, United Bible Societies.

Hatim, B. (2001): *Teaching and Researching Translation,* Harlow and New York, Longman.

Kang, B.-M. and H-G. Kim (2000): *Hangukeoeui tekst jangreu mwunche yuhyeong*: *keompyuteowa tonggyejeok gibeopeui iyong* [Text Genre, Style, Types in Korean Language: Application of Computer and Statistical Methods], Seoul, Taehaksa.

Klaudy, K. (1998): "Explicitation," *Routledge Encyclopedia of Translation Studies* (M. Baker and K. Malmkjær, eds.), London and New York, Routledge, p. 80-85.

LAVIOSA, S. (1997): "How Comparable Can Comparable Corpus Be?," *Target* 9-2, p. 289-319.

LEPPIHALME, R. (1997): *Culture Bumps*: *An Empirical Approach to the Translation of Allusions,* Clevedon and Philadelphia, Multilingual Matters.

MUNDAY, J. (1998): "A Computer-assisted Approach to the Analysis of Translation Shifts," *Meta* 43-4, retrieved 2 Feb. 2003, <http://www.erudit.org/revue/meta/1998/v43/n4> .

NAM, K.-S. and Y.-G. Ko ([1985] 1993): *Pyo-jun-kuk-eo-mun-beop-ron* [Standard Korean grammar], (revised edition), Seoul, Tap.

NIDA, E. (1964): *Toward a Science of Translating*: *With Special Reference to Principles and Procedures Involved in Bible,* Leiden, E.J. Brill.

NIDA, E. and C. Taber (1969): *The Theory and Practice of Translation,* Leiden, E.J. Brill.

NORD, C. (1991): *Text Analysis in Translation,* Amsterdam, Rodopi.

—— (1996): "Revisiting the Classics: Text Type and Translation Method – An Objective Approach to Translation Criticism," *The Translator* 2-1, p. 81-88.

OLOHAN, M. (2002): "Comparable Corpora in Translation Research: Overview of Recent Analyses Using the Translational English Corpus," *LREC Language Resources in Translation Work and Research Workshop Proceedings*, p. 5-9.

ØVERÅS, L. (1998): "In Search of the Third Code: An Investigation of Norms in Literary Translation," *Meta* 43-4, <http: //www.erudit.org/revue/meta/1998/v43/n4>.

SOHN, H-M. and J. HAIG (1997): *Japanese/Korean Linguistics,* John Haig Center for the Study of Language and Information.

SPERBER, D. and D. Wilson (1986): *Relevance*: *Communication and Cognition,* Oxford, Blackwell.

TANNEN, D. (Ed.). (1993a): *Framing in Discourse,* New York and Oxford, Oxford Univ. Press.

—— (1993b): "What's in a Frame? Surface Evidence for Underlying Expectations," *Framing in discourse* (D. Tannen, ed.), New York and Oxford, Oxford Univ. Press, p. 14-56.

TOURY, G. (1995): *Descriptive Translation Studies and Beyond,* Amsterdam and Philadelphia, John Benjamins.

VANDERAUWERA, R. (1985): *Dutch Novels Translated into English: The Transformation of a "Minority" Literature,* Amsterdam, Rodopi.

VERMEER, H. (1989): "Skopos and Commission in Translation Action," *Readings in Translation Theory* (A. CHESTERMAN, ed.), Helsinki, Oy Finn Lectura Ab.