CORPUS LINGUISTICS

Another, recent but fast-growing field of linguistics that has already provided translation theorists with valuable information, is that of corpus linguistics. The study of language on the basis of text corpora can be traced back to around 1960 with the launch of the Survey of English Usage (SEU) at London University and the advent of computers which made it possible to store large amounts of material. The first machine-readable corpus compiled at Brown University in the early 1960s was soon followed by others such as the London-Oslo/Bergen (LOB) Corpus. Capitalising on the combined strengths of the Brown and SEU corpora, starting in 1975, Jan Svartvik and his colleagues at Lund University in Sweden, rendered the unscripted spoken texts of the SEU corpus machine-readable. This resulted in the London–Lund Corpus (LLC), an unmatched resource for the study of spoken English. While the Brown Corpus and the LOB Corpus may have seemed vast at the time, their size has been easily surpassed as massive amounts of machine-readable texts have become available as a by-product of modern electronic communication systems.

Since its beginning in the 1960s, the corpus as a source of systematically retrievable data, and as a test bed for hypotheses, has become widely used by linguists, resulting in findings that include some with obvious implications for translation. Using a corpus consisting of 75 novels published 1967-1977, half of which were novels originally written in Swedish and half were translations, Martin Gellerstam of the University of Gothenburg has systematically compared original texts with texts in translation. His early 1986 study as well as later ones point squarely to the influence in translation of the source on the target text and also revealed previously-unobserved cross-linguistic differences between the two languages involved in the translation process. Another early corpus study of the influence of English on lexical selection in Danish confirmed Gellerstam's findings, showing that modality, typically expressed in many Germanic languages through the use of modal particles (such as jo' and 'vel' in Danish), was greatly underrepresented in texts in translation. In their place instead appeared English-influenced modal verb constructions such as 'I presume' or 'I suppose', often resulting in a marked awkwardness of style. More recently, translation corpora have investigated the fate of another kind of particle in translation, so-called discourse particles such as 'oh', 'well' and 'now', which tend to express emotional attitudes and contribute towards the coherence of the utterance. In translation into other languages, these discourse particles are frequently rendered in a multitude of different ways and the availability of translation corpora now makes possible a study of the semantic and contextual reasons underlying the translator's choice as shown by Aijmer drawing on an English/Swedish Parallel Corpus. Other language pairs for which parallel corpora have been compiled include English/French, English/Italian, English/Norwegian and English/German.

- Anderman, Gunilla. "Chapter 3. Linguistics and Translation". ACompanion to Translation Studies, edited by Piotr Kuhiwczak and Karin Littau, Bristol, Blue Ridge Summit: Multilingual Matters, 2007, pp. 45-62. https://doi.org/10.21832/9781853599583-005

1. What is corpus linguistics?

A) The study of linguistic data based on text corpora

B) The analysis of cross-linguistic differences in translation

C) The use of electronic communication systems for language learning

D) The compilation of machine-readable texts for linguistic research

2. When did the study of language on the basis of text corpora begin?

A) Around 1970s

B) In the early 1960s

C) In the late 1960s

D) In the 1960

3. What is the London–Lund Corpus?

A) A compilation of machine-readable texts from the 1960s

B) A machine-readable corpus of spoken English

C) A test bed for hypotheses in corpus linguistics

D) A resource for the study of cross-linguistic differences

4.What did Martin Gellerstam's study reveal about translations?

A) The influence of the source language on the target text

B) The underrepresentation of modality in translated texts

C) The importance of lexical selection in translation

D) The significance of discourse particles in translation

5. What technological development enabled the storage of large amounts of material for corpus linguistics?

A) Launch of the Survey of English Usage (SEU)

B) Advent of computers

C) Compilation of the Brown Corpus

D) Availability of machine-readable texts

Answer 1: A) The study of linguistic data based on text corpora

Answer 2: D) Around 1960

Answer 3: B) A machine-readable corpus of spoken English

Answer 4: A) The influence of the source language on the target text

Answer 5: B) Advent of computers