A classification model for translational correspondences

Martha Thunes

University of Bergen

martha.thunes@lle.uib.no

- Thunes, Martha. 2011. *Complexity in Translation. An English-Norwegian Study of Two Text Types.* Doctoral dissertation. University of Bergen.
- Based on a study of pieces of translationally parallel texts of English and Norwegian.
- Two text types: law text and fiction.
- Focus here: model for analysing translationally corresponding text units.

Research questions

- 1. To what extent is it possible to compute the actual translation relation found in the selected English-Norwegian parallel texts?

- Assume: fully automatic machine translation (MT), based on linguistic information.

- 2. Is there, in the empirical material, a difference in the degree of translational complexity between the two text types, law and fiction?

- Restricted text type: law. Unrestricted: fiction.

A measurement of translational complexity

- Manual analysis of running parallel texts.

- Extraction of pairs of translationally corresponding units, string pairs.

- Primary unit of translation: the finite clause.

- Each string pair has been analysed according to the model: "the correspondence type hierarchy".

The correspondence type hierarchy

- Original source: Helge Dyvik, University of Bergen.
- 4 main categories of translational correspondences.
- To be illustrated with examples from:

  Vik, Bjørg. 1979. *En håndfull lengsel.* Oslo: J. W. Cappelens Forlag AS.

  Vik, Bjørg. 1983. *Out of Season and Other Stories.* Translated by David McDuff and Patrick Browne. London: Sinclair Browne.

Type 1

(1a) Hun har vært en skjønnhet.

'She has been a beauty.'

(1b) She has been a beauty.

- Target sentence corresponds word-by-word with the source sentence.

- Relatively infrequent in the language pair English-Norwegian.

Type 2

- The sequence of constituents may differ:

(2a)   Dessuten virket hun overlegen.

      'Also looked she haughty.'

(2b) She also looked haughty.

- Differences w.r.t. grammatical form words:

(3a)   Leiligheten var ufattelig rotete.

      'Flat.DEF was unbelievably untidy.'

(3b) The flat was unbelievably untidy.

Type 2, ctd.

- Every lexical word in the ST has a correspondent in the TT of the same lexical category and with the same syntactic function as the source word.

- This type, too, is relatively infrequent in the language pair English-Norwegian.

Type 3

- At least one structural difference violating equivalence between ST and TT with respect to syntactic categories and functions.

- No mismatch between original and translation on the semantic level.

- In this language pair type 3 correspondences seem to be more frequent than each of the two lower types.

Type 3, ctd.

(4a) Hildegun himlet lidende mot taket og svarte med uforskammet høflighet.

'Hildegun rolled-eyes suffering towards ceiling.DEF and answered with brazen politeness'

(4b) Hildegun rolled her eyes in suffering towards the ceiling and answered with brazen politeness.

Type 4

- Differences between ST and TT not only on the structural level, but also on the semantic.
- The most frequent correspondence type in this study.

Type 4, ctd.

(5a) Her kunne de snakke sammen uten å bli ropt inn for å gå i melkebutikken eller til bakeren.

'Here could they talk together without to be called in for to go in milk-shop.DEF or to baker.DEF'

(5b) They could talk here without being called in to go and buy milk or bread.

Subtypes

- A set of semantic subcategories within the main correspondence types 3 and 4.

- Recurrent phenomena involving some kind of semantic deviation between translationally corresponding units.

- Classes of phenomena: subtypes within the main correspondence types 3 and 4.

- Cf. chapter 6 in Thunes (2011).

Product-oriented analysis

- The correspondence types are not translation methods or strategies.

- They are descriptions of correspondence relations between linguistic expressions.

- I do not ask: "what kinds of knowledge has the translator used in order to produce a given TT?"

- But: "what kinds of information about ST expressions are necessary in order to produce the chosen target expressions?"

The translational relation

- *La langue*, the language system seen in abstraction from actual language use, versus *la parole*, the language when used as a means of communication.
- Dyvik (1998: 51-52):
  The translational relation between a source text and its translation: on the level of *parole*.
  The translational relation between two languages: on the level of *langue*.

The translational relation, ctd.

- Different factors determine, respectively, the *parole* relation and the *langue* relation.
- Whereas the *langue* relation is determined by information about SL and TL and their interrelations, translational correspondences between situated units of text may be determined also by additional information sources —
  e.g. information from the context of the source expression, information about the purpose of the ST, about the purpose of translating it, and other kinds of background information.

Predictability/computability

- There is a translational relation between the inventories of simple and complex linguistic signs in two languages which is predictable, and then also computable, from information about SL and TL, and about how the language systems correspond.

- Computable translations are predictable from the linguistic information coded in the ST, together with information about SL ⇄ TL.

Predictability/computability (ctd.)

- Non-computable translation tasks require access to additional information sources, e.g. general or task-specific extra-linguistic information, or task-specific linguistic information from the context surrounding the source expression.

- Types 1–3: the domain of linguistically predictable, or computable, correspondences.

- Type 4: the domain of non-predictable, or non-computable, correspondences.

Linguistically predictable correspondences

- Cf. Dyvik (1998, 1999):

- To identify the translational relation between the sign inventories of two language, the *langue* relation, is to find the linguistically predictable correspondences of that language pair.

- Such sign correspondences hold between signs with shared meaning properties.

Linguistically predictable correspondences, ctd.

- The criterion of shared meaning properties specifies what must *at least* be present in a linguistically predictable correspondence.

- Linguistically predictable translations do not necessarily share *all* the meaning properties associated with a given SL sign.

- Language systems are differently structured in terms of grammar and lexical inventory.

Linguistically predictable correspondences, ctd.

- The linguistically predictable translations of a given sign in the SL is the set of signs in TL which exhibit a maximum of the meaning properties of the SL sign, having taken into account differences in semantic structure between the two languages.

The correspondence type hierarchy

- From type 1 upwards it reflects an increase in the degree of translational complexity.

- This increase is correlated with an increase in the types and amounts of information needed to solve given translation tasks.

- Translational complexity: chapter 3 in Thunes (2011).

The empirical investigation

- Requires a human, bilingually competent analyst.

- This study: the data comprise about 68 000 words.

- 2 pairs of law text, 4 pairs of fiction texts.

- Both directions of translation.

Measuring translational complexity

- Calculate the distribution of the four types of translational correspondence within the set of string pairs extracted from a body of parallel texts.

- The most important aspect shown by the empirical results is the division between computable and non-computable correspondences.

- Computable or not: is there at least one linguistically non-predictable semantic deviation between source and target string?

Results

- Across the entire data set, calculated in terms of string lengths:

  types 1+2+3 cover 44,8%

  type 4 covers 55,2%

- Within the law data types 1+2+3 cover 50,2%.

- Within fiction types 1+2+3 cover 39,6%.

- The empirical material is small; no generalisations.

Too pessimistic?

- Are the analysed parallel texts an appropriate standard for automatisation?

- They represent an ideal for the end result, and not for the raw output of an MT application.

- High-quality translation without post-editing, or revision, is uncommon also when the translator is human.

Automatic translation fruitful?

- The workload potentially involved in correcting machine output?

- A higher frequency of non-predictable semantic differences in the relatively unrestricted fiction texts than in the restricted law texts.

- The frequency of string pairs with only one minimal semantic deviation is considerably higher in law than in fiction.

Nonfinite-finite

- Correspondences between English nonfinite constructions and Norwegian finite subclauses constitute the most frequent type of minimal semantic deviation.

(6a) de områder som er nevnt i tillatelsen

   'the areas which are mentioned in license.DEF'

(6b) the areas mentioned in the licence

- Noticeably frequent among the law text data.

Tentative conclusions

- In the investigated pairs of law texts, MT may be helpful, provided that the effort required by post-editing is smaller than that of manual translation.

- In the fiction texts, post-editing of automatically generated translations would be laborious and not cost effective.

Comments on the classification model

- Requires manual analysis by language expert.
- Limited amounts of text; scaling up is difficult.
- Language-pair independent.
- So far: English-Norwegian (Hasselgård 1996, Thunes 1998, 2011); English-Portuguese (Tucunduva 2007, Silva 2008, Azevedo in progress).
- Also a method for contrastive analysis.

Comments on the classification model, ctd.

- Hasselgård (1996): too coarse-grained for contrastive analysis.

- Thunes (1998, 2011): semantic subtypes.

- Tucunduva (2007), Silva (2008), Azevedo (in progress): the model offers a consistent way of describing all pairs of translationally matched units in a given body of parallel text.

- A simple and flexible model.

References

Aijmer, Karin, Bengt Altenberg, and Mats Johansson (eds). 1996. *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies, Lund 4–5 March 1994. Lund Studies in English* 88. Lund: Lund University Press.

Azevedo, Flávia. In progress. *Investigating the problem of codifying linguistic knowledge in two translations of Shakespeare's sonnets: a corpus-based study.* Doctoral dissertation. Federal University of Santa Catarina, Florianópolis.

Dyvik, Helge. 1998. A translational basis for semantics. In: Johansson and Oksefjell (eds), 1998, 51–86.

Dyvik, Helge. 1999. On the complexity of translation. In: Hasselgård and Oksefjell (eds), 1999, 215–230.

Hasselgård, Hilde. 1996. Some methodological issues in a contrastive study of word order in English and Norwegian. In: Aijmer et al. (eds), 1996, 113–126.

Hasselgård, Hilde and Signe Oksefjell (eds). 1999. *Out of Corpora. Studies in Honour of Stig Johansson. Language and Computers: Studies in Practical Linguistics* 26. Am-sterdam and Atlanta, GA: Rodopi.

Johansson, Stig and Signe Oksefjell (eds). 1998. *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*. *Language and Computers: Studies in Practical Linguistics* 24. Amsterdam and Atlanta, GA: Rodopi.

Silva, Norma Andrade da. 2008. *Análise da tradução do item lexical evidence para o por-tuguês com base em um corpus jurídico.* Master's thesis. Federal University of Santa Catarina, Florianópolis.

Thunes, Martha. 1998. Classifying translational correspondences. In: Johansson and Okse-fjell (eds), 1998, 25–50.

Thunes, Martha. 2011. *Complexity in Translation. An English-Norwegian Study of Two Text Types.* Doctoral dissertation. University of Bergen.

Tucunduva, Camila de Andrade. 2007. *Translating completeness: a corpus-based approach.* Master's thesis. Federal University of Santa Catarina, Florianópolis.