On semantic differences between translated and non-translated Dutch. Using bidirectional parallel corpus data for measuring and visualizing distances between lexemes in the semantic field of inceptiveness

Lore Vandevoorde, Gert De Sutter & Koen Plevoets

1. Introduction

In her 1993 seminal paper "Corpus Linguistics and Translation Studies," Mona Baker proposed a research programme for translation universals (also: TU's) which she defined as "features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems" (Baker, 1993: 243). Translated texts were thought to be more explicit (TU of explicitation) and simpler (TU of simplification) than original language, the overall language use in translated text more conservative (TU of conservatism). This opened the way for numerous corpus-based studies concerned with both the validation and refutation of the so-called universals which have been operationalized via different linguistic features (Malmkjaer 1997; Laviosa 1998, 2002; Mauranen 2000; Olohan & Baker 2000; Baker 2004; Bernardini & Ferraresi 2011, Delaere et al. 2012; De Sutter et al. 2012 – see Kruger 2012 for an overview).

Given the attested linguistic differences between translated and non-translated language, one could wonder whether differences on the semantic level exist too. Translational features on the *semantic level* have though been somewhat neglected (Laviosa, 2002: 28). The question, for instance, whether semantic relations between words in a specific semantic field are identical in translated and original language has rarely been raised[1] within translation studies. Unsurprisingly though, as strategies to detect simpler, more explicit or more conservative language (via the comparison of grammatical structures or vocabulary between translated and original texts) do not necessarily apply to semantic networks. How can we recognize a 'more conservative' semantic relation, how can we even see it (in contrast)? Admittedly, Cognitive Translation Studies have engaged with this question, but so far, the discussion has been pursued mainly on a theoretical level. In that respect, models of bilingual semantic representation have been proposed by researchers like Halverson (2003, 2010), combining psycholinguistic models of bilingual semantic representation and cognitive-linguistic concepts, like the salience of prototypes and network schemas. Halverson affirms that these models of bilingual semantic representation can help us to understand the workings of translational phenomena and she advocates the use of combined experimental and corpus-based methodologies to understand the patterns found in parallel corpora from a cognitive perspective. Nevertheless, the step towards descriptive testing has not yet been taken.

Hence, the aim of this paper is to make a first attempt towards measuring semantic differences between translated and non-translated language. More particularly, we present a quantitative bottom-up corpus-based method for the identification of lexical items in a semantic field. The proposed method will enable us to measure and to visualize semantic similarity between the elements in that field (i.c. the field of inceptiveness in Dutch), using bidirectional parallel corpus data (Dutch-French). This method builds on the successful implementation of parallel corpora within contrastive linguistics to discern semantic fields (Dyvik 1998; 2004; Aijmer & Simon-Vandenbergen 2004, 2006; Simon-Vandenbergen 2013), while simultaneously overcoming one of its drawbacks, viz. the accurate, statistics-based visualization of the observed fields.

The structure of this paper is as follows: In section 2, an overview is given of the way parallel corpora have been used recently by contrastive corpus linguists for the investigation of semantic issues. Then, we put forward a translational approach to the retrieval of semantic fields based on a technique of back-and-forth translation (section 3). Next, semantic similarity is measured via the statistical technique of correspondence analysis (section 4), which enables us to visualize the semantic fields of translated and original Belgian-Dutch inceptiveness. Finally, the last section summarizes the main findings of this study and looks ahead to future research steps.

2. Background

Whereas the empirical study of semantic differences in Corpus-based Translation Studies is still in its infancy, contrastive linguists have successfully developed and used methods based on translational equivalence[2] to define semantic properties of and relations among lexemes, providing thus an empirical basis for semantic claims (Noël, 2003). The underlying idea is that cross-linguistic lexicalization can determine the different senses of a word (Resnik and Yarowsky, 1997, 1999): "[…] if another language lexicalizes a word in two or more ways, there must be a conceptual motivation" (Ide et al., 2002: 61). A method that is well known in this regard is Dyvik's Semantic Mirrors approach.

Based on the assumption that semantically closely related words ought to have strongly overlapping sets of translations, Dyvik (1998, 2004) purports the use of parallel corpora for the identification of semantic relations. In his own research, he uses parallel corpora to derive large-scale semantically classified vocabularies for machine translation and other kinds of multilingual processing (1998: 51). His Semantic Mirroring Technique also gives way to a translational basis for semantic descriptions in a context that is wider than computational linguistics. He attributes several advantages to the use of parallel corpora, for translation is both a "large-scale" and a "normal kind of" linguistic activity that does not involve any kind of "meta-linguistic, philosophical or theoretical reflection" (Ibid.), a property rather difficult to obtain in (contrastive) linguistic studies. Dyvik's methodology has been acknowledged for its ability "to define lexical properties as ambiguity, vagueness and synonymy, as well as lexical fields, feature-specified hierarchies and overlap relations with these fields (e.g. prototypicality, hyponymy)" (Altenberg & Granger, 2002: 29).

Several researchers have made use of (a derived form of) the Semantic Mirrors, mostly for intralinguistic and contrastive purposes, and with respect to discourse markers (Mortier & Degand, 2009), pragmatic markers (Aijmer & Simon-Vandenbergen 2004), and adverbs (Simon-Vandenbergen 2013). Aijmer and Simon-Vandenbergen (2004), for instance, describe the semantic field of expectation in English by looking at the Dutch and Swedish translations of English lexemes. The sets of translations back into English (called the 'back-translations') of both the Dutch and the Swedish translations are then compared to each other cross-linguistically. From their study, it appears that similar back-translations around different pivot languages do indeed indicate similar meanings. Aijmer and Simon-Vandenbergen also point out some advantages of a method based on parallel corpora:

> Firstly, the translation data can be used for a more detailed description of the polysemy of a lexical item […]. Secondly, the picture emerging from the approach shows which meanings are close to each other and which are distant or peripheral in the field. Thus the translations are in many ways more reliable than the paraphrases provided by earlier researchers and can confirm or reject meaning hypotheses based on a single language only. (Aijmer & Simon-Vandenbergen, 2004: 1786).

They conclude that mirroring allows for an expansion of the semantic field with lexemes that intuitively would not appear in the described semantic field. The (simplified) representations of the semantic fields yielded via mirroring provide information about the semantic fields in all the languages involved in the cross-linguistic comparison (Aijmer & Simon-Vandenbergen, 2004: 1797).

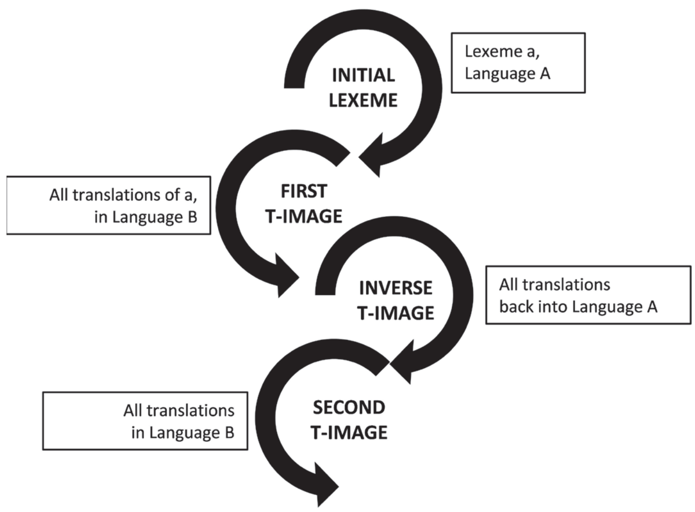
3. Methodology

In this paper, we apply Dyvik's Semantic Mirroring technique, which uses Ivir's (1983, 1987) procedure of back-translation to control for unwanted translational effects, such as translators' idiosyncrasies or particular communicative or textual strategies applied in translation (Altenberg & Granger 2002: 17). One important difference with previous applications of the Semantic Mirroring is that we apply Dyvik's technique in such a way that (i) it creates semantic fields of inceptiveness in both translated and original Dutch[3] and (ii) it provides a statistically reliable way of visualizing semantic distances between the lexical items in the semantic fields. In this section, we first explicate the Semantic Mirroring Technique (section 3.1) as it was described by Dyvik. Then, we propose an extension of the technique in order to create both translated and original semantic fields (section 3.2). Finally, we apply the extended technique to the semantic field of Dutch inceptiveness (3.3).

3.1 Semantic mirrors

Dyvik starts from an initial polysemous *lexeme a in Language A* and extracts all its translations in Language B manually from the English-Norwegian Parallel Corpus (ENPC), a sentence-aligned corpus. He calls this set of translations the *first T-image of a in Language B.* Then, commensurably, the translations back in Language A (the back-translations) of the T-image (themselves translations from a) are looked up. This is called the *Inverse T-image of a in Language A.* It is worth noting that at this point, Dyvik's method differs considerably from classical translation-based WSD techniques, where translations are merely used to disambiguate between the senses of the initial lexeme (e.g., Ide 2002; Lefever 2012). Finally, the initial procedure is applied a second time: the translations in Language B of the Inverse T-image lexemes in Language A are looked up (this is called the *second T-image*), resulting in a structure that depicts the senses of both Language A and Language B lexemes. Schematically, the procedure looks as follows:

Schematic representation of Dyvik's Semantic Mirroring Technique



Applied to the example of the Dutch polysemous lexeme *papier* (lexeme a in Language A), we obtain a T-image in English (Language B) of *papier* with *paper, sheet, document, bond.*

The Inverse T-image back into Dutch (Language A) of these English (Language B) lexemes then looks as follows:

[paper] *papier, krant, toets, behangpapier, bankbiljet;*
[sheet] *laken, blad;*
[document] *document, akte;*
[bond] *obligatie.*

The Inverse T-image allows us to differentiate between Dutch papier$_1$ (the material), papier$_2$ (a blank sheet of paper) and papier$_3$ (a valuable piece of paper). Finally, the resulting Dutch lexemes of the Inverse T-image are used to create a second T-image in English, allowing a structuration of sense relations in both Language A and B. Hereunder follows the second T-image. The underlined lexemes are the ones recurring from the first T-image:

[papier] paper, sheet, document, bond
[krant] paper, *newspaper, daily;*
[toets] *test,* paper, *analysis, key;*
[behangpapier] *wall*paper;
[bankbiljet] *note, bill,* paper *currency*
[laken] sheet, *tablecloth;*
[blad] *leaf, tray,* sheet, paper;
[document] *document,* paper;
[akte] document, *contract, act, contract, deed;*
[obligatie] bond, *debenture.*

3.2 Extending the Semantic Mirroring Technique

3.2.1 A rationale for extension

We now implement two new elements in Dyvik's technique in order to extend its use for the creation of translated and original semantic fields. First, drawing on Dyvik's (2004: 311) assumption that "semantically closely related words ought to have strongly overlapping sets of translations," overlapping sets of translations should commensurably reveal semantic relations between translations, between translations and their source language items and between the source language items themselves. Dyvik uses translations as a means to reveal semantic relations, but does not implement the nature (translated or original) of the data that are used at the different stages of the Mirroring. However, within Semantic Mirrors, Language A is a source language in the first and the second T-image and a target language in the Inverse T-image. This implies the possibility to distinguish between translated and original Dutch semantic fields within the Mirroring Technique. Another difficulty is that the Mirroring Technique does in fact only reveal the relational structure between polysemous lexemes, but does not tell us anything about the degree of semantic similarity between the lexemes in the created field. The degree of similarity can be implemented by inserting (source and target language) frequencies into the rationale. This will enable us to measure and visualize semantic similarities between the lexical items in the semantic field of inceptiveness, once as a 'translated' semantic field and a second time as an 'original' semantic field. By doing so, we use the Semantic Mirroring Technique in such a way that is useful for Corpus-based Translation Scholars who are interested in semantic differences between translations and original texts. Moreover, we use the (quantitative) output of the Mirroring Technique and plug it in the statistical technique of correspondence analysis in order to adequately measure and visualize the semantic relationships between the lexical items in the semantic field.

3.2.2 Applying the extended Mirroring technique

The extended technique works as follows. First, all translations of a given (set of) lexeme(s) in a large parallel corpus are checked manually (T-image). Then, inversely, all translations of these translations back into the initial source language (Dutch) are looked up in the same parallel corpus. This is the Inverse T-image. These 'back-translations' enable us to access the structure of the semantic field via the first-order translations. Via an application of the statistical technique of correspondence analysis, this leads to the first visualization, which includes the lexemes of the *Inverse T-image* of an initial set of lexemes. All members of the *Inverse T-image* are in fact translations (of the lexemes of the first T-image). Their exact position in the semantic field and the distance to other lexemes in that field is based on their frequencies as translations of the first T-image and thus depicts their position in the semantic field in translated Dutch via a statistically founded visualization. We can consider the created semantic field as a translated semantic field, for both the variation and the frequencies of the lexemes are determined on the basis of translation.

In order to create a non-translated semantic field comparable to the translated semantic field, we take the set of lexemes of the previous Inverse T-image but we now consider these lexemes as *source-language items*. This means that (i) we take the Dutch lexemes of the previous Inverse T-image, (ii) we query them from the parallel corpus as source language items (this corresponds to Dyvik's second T-image), (iii) we look up their translations in the corpus and (iv) we visualize the Dutch lexemes (as source-language items) via the same statistical technique of correspondence analysis. In this second visualization, the position of the lexemes is based on their frequencies as source language items, and, in this way, it depicts the semantic field in original Dutch. Translational data have thus been used solely as a sense-discrimination technique. The plotted lexemes themselves are the same in both visualizations/plots, which makes them comparable. This technique enables us to regard the first visualization as a representation of the *translated* semantic field of inceptiveness and the second visualization as a representation of the *original* (non-translated) field of inceptiveness.

3.3 Applying the method to the Dutch semantic field of inceptiveness

The data for this study were extracted from the Dutch Parallel Corpus (DPC), a 10-million-word, sentence aligned, both parallel and comparable corpus. It is balanced with respect to five text types (external communication, journalistic texts, instructive texts, administrative text, fictional and non-fictional literature) and four translation directions (Dutch to French, French to Dutch, Dutch to English and English to Dutch). Each text type accounts for 2,000,000 words and within each text type, each translation direction contains 500,000 words (Macken et al., 2011: 376–378). Due to copyright difficulties (a persistent obstacle in large corpus building including fictional texts), the DPC is not balanced for fictional literary texts and so the results of our study do not apply for this text type. We chose to extract only the Belgian-Dutch data from the DPC. We did not take into consideration the data for Dutch (Netherlandic)-Dutch. The text providers of the DPC are, save a few, all Belgian (Netherlandic Dutch providers supplied mostly fictional literature, a genre that cannot be taken into consideration for this study due to the scarcity of the data), hence our choice to eliminate the Dutch-Dutch data.

In order to generate the semantic field of BEGINNEN and to initiate the technique of back-and-forth translation, we selected a concise set of near-synonyms of BEGINNEN "to begin," consisting of *aanvangen, een aanvang nemen, starten, van start gaan* and *aanvatten*. Our selection is based on careful lexicographic analysis, starting from the most prototypical expression of the concept of inceptiveness, namely BEGINNEN [TO BEGIN]. First, we examined eight dictionaries[4] for their synonyms of *beginnen* without taking into account any lexicographic meta-information about hypernymy or hyponymy[5] provided by the dictionaries (for different dictionaries have different policies about meta-data). Nineteen out of a total of 104 synonyms were attested in at least three of the eight consulted dictionaries. After having extracted all sentences in the DPC-corpus containing BEGINNEN (n = 1,435), we subjected all these sentences to an interchangeability test for these 19 lexicography-based synonyms. This resulted in a set of five prototypical synonyms.

The French translations of this set of onomasiological variants of BEGINNEN (n = 564) were manually checked,[6] returning a total of 74 different translations. We then selected those French lexemes that were attested as translations of at least two of the initial Dutch lexemes (minimal overlap criterion). Furthermore, we applied a frequency threshold of 5, signifying that the French types had to be attested at least five times in the corpus as translations of the initial set of Dutch lexemes. This yielded a T-image of 12 different French lexemes. The T-image lexemes were inversely queried from the corpus as source-language lexemes (n = 1,064). Their translations back into Dutch were manually checked, and, applying the same selection criteria (minimal overlap, frequency threshold of 5), this resulted in an inverse T-image of 22 Dutch lexemes. The resulting frequency tables of both the T-image (Dutch rows, French columns) and the inverse T-image (French rows, Dutch columns) were analysed with the technique of correspondence analysis (Greenacre, 2007; Lebart et al., 1998). Correspondence analysis arrives at a lower-dimensional representation of the row and column associations, thereby visualizing the semantic distances between the lexical items in the field. The position of these 22 lexemes is based on their frequencies as translations and thus depicts their position in the semantic field of inceptiveness in translated Dutch via a statistically founded visualization.

In order to visualize the original semantic field, the 22 Dutch lexemes were now inversely queried from the corpus as source-language lexemes (n = 5,322) and their French translations were checked (second T-image).

The visualization of the original semantic field uses the same selection of lexemes and the same statistical technique as the visualization of the translated semantic field, which makes them comparable, but differs in this way that it is based on source-language frequencies, which ensures their visualization as original Dutch. The use of translational data for the second visualization is limited to a sense-discrimination technique and has no further impact on their position in the semantic field.

4. Results: measuring and visualizing semantic similarity between lexical items

4.1 Correspondence analysis

Figure 1 shows the semantic field of translated Dutch BEGINNEN. We observe that most lexemes (and in fact all lexemes that were selected after the lexicographic analysis) are in the plot's origin, viz. *beginnen* [to begin], *aanvangen* [to start], *van start gaan* [to start], *starten* [to start], *aanvatten* [to commence]. This cluster can consequently be interpreted as the prototypical centre, consisting of lexemes with the basic meaning of the inceptive category, viz. "start of a general process." The lexemes of the second cluster have in common that none of them are verbs. This would imply that, for instance, *aanvankelijk* [first] and *begin* [beginning] are semantically closer to each other than *begin* [beginning] is to *beginnen* [to begin]. We also observe two outlying lexemes: *openen* [to open] and *vertrekken* [to leave] and a small outlying cluster with *invoeren* [to introduce] and *instellen* [to establish]. *Invoeren* and *instellen,* mostly refer to a "rule or legislation becoming effective," so they typically appear in formal, legislative texts, hence their outlying position. The inchoative meaning of *openen* is (i) a formal form of inchoativity, as in "to open a sitting or a meeting," and (ii) a metaphor (as in: "His new job *opened doors* for his future"), which could explain its outlying position. Finally, the outlying position of *vertrekken* [to leave] in the translated semantic field could be due to translational interference. The position of the Dutch lexemes in the translated field depends on the underlying position of the French lexemes, so their position can actually inform us on the (possible) translational effects. The closest neighbour lexemes of *vertrekken* in the plot are the cluster of non-verbs, for instance *aanvankelijk* [first, in the beginning] and *aanvang* [outset]. This leads to the hypothesis that *vertrekken* in its gerund form *vertrekkende (van)* [departing from] is semantically closer to *aanvankelijk* and *aanvang* than to the prototypical inceptive lexemes. This gerund form could be an example of translational *shining through* of French *à partir de* [leaving from, as from].

Now we take a look at Figure 2 (the semantic field of original Dutch BEGINNEN). The most obvious resemblances between the translated and the original semantic field are the two main clusters: in the origin of the original Dutch plot, we find the prototypical centre and we equally observe a separate cluster with the non-verbs. The lexemes in each of the clusters are almost identical to the ones of the translated plot (apart from *vertrekken*, which is now in the plot's origin). We do notice that, in the origin, lexemes are further apart from each other, which indicates that small meaning differences are somewhat flattened in translated language. Furthermore, we observe that *vertrekken* has become prototypical. Based on this observation, we could conclude that *vertrekken* is a prototypical expression of inceptiveness in original Dutch, but it is not used as such in translated Dutch, hence its outlying position. *Openen* remains its outlying position, which confirms the findings in the translated plot. Overall, we did not observe any major differences between the semantic fields of translated and original Dutch, although we did detect some smaller differences, assumedly pointing out differences between the original and the translated semantic field.

4.2 Correspondence analysis with anchoring

From the visualizations of the first correspondence analysis (see Figures 1 and 2), we observed that lexemes with the same word class seemed to systematically cluster together, an outcome one would not intuitively expect (we would not expect *aanvankelijk* [first] and *begin* [beginning] to be semantically closer to each other than *begin* [beginning] to *beginnen* [to begin]). If, in fact what we want to know is how the lexemes are related within the semantic field of inceptiveness, it might be desirable to first restrict the semantic space in which the lexemes will appear to the field of inceptiveness. For the second correspondence analysis (see Figures 3 and 4), we thus decided to first create a "space of inceptiveness" based on the first T-image. More specifically, before we actually plotted the 22 Dutch lexemes, we applied a correspondence analysis on the original six Dutch lexemes with their 12 French translations. This had the effect that the positions of the 12 French lexemes were 'anchored' with respect to the six prototypical

Dutch lexemes. In other words, this allowed us to create a 'stable' space of inceptiveness. On the basis of these 12 'anchored' positions, the 22 Dutch lexemes were subsequently visualised,[7] once as target-language items (translated Dutch, Figure 3) and once as source-language items (original Dutch, Figure 4).

When we look at these analyses (Figures 3 and 4), we find a different distribution. In both Figures 3 and 4, we find *beginnen* in the plot's origin. So far, this is the same observation as with the first correspondence analysis. But when we look at the separate clusters in each figure, we find that they are now clearly meaning-based and formed independently of the word class of each lexeme. This could be explained by the fact that the position of the lexemes in this second analysis is clearly restricted to the field of inceptiveness (by the previous creation of the 'stable inceptive space'), which avoids the position of the lexemes to be biased by their relation to other semantic fields than the one of inceptiveness.

In Figure 3 (translated semantic field) we find, next to the prototypical centre, a second cluster with lexemes like *oprichten* [to set up], *lanceren* [to launch], *opstarten* [to start up], all referring to the "beginning of a project, an initiative or a business." *Invoeren* [to establish] and *instellen* [to set up] are again outlying, which confirms our analysis based on classical statistical techniques. *Aanvangen* [to commence] and especially *ingaan* [to take effect] are outlying, which could be due to their formal character. Figure 4 shows us the original Dutch semantic field. Parallel to the classical analysis, we observe that the lexemes are lying further apart, which shows that the differences between the lexemes are more clearly expressed. This confirms our idea from the first analysis that translation flattens meaning differentiation. We again notice two clusters, the one in the origin is the prototypical one, the one to the right of the prototypical cluster consists of lexemes referring to the "beginning of a project, an initiative or a business." Note that in the original Dutch field, the lexemes seem to gradually descend from the protypical centre, towards the right, and towards a slightly outlying position. Following this line from centre to periphery, we clearly remark that the lexemes become more formal, with at the end of this line, *invoeren* [to establish] and *instellen* [to set up]. Note also the difference with the translated semantic field where those two lexemes are clearly outlying. This shows that the gradual meaning differentiation we observe in original language has somewhat disappeared in translated language. Also parallel to the translated semantic field, we observe *aanvangen* [to commence]. Both in translated and in original language, this lexeme seems to hold a kind of middle position, as shown by its similar position in both plots. Finally, we see that outlying *ingaan* [to take effect] is now rejoined by *vanaf* [as from]. In original Dutch, the inceptive aspect of *vanaf* is rather remote, whereas in translated language, *vanaf* is even prototypical. This could again be explained via translation: *vanaf* is often a good equivalent for many of the French inceptive lexemes like *débuter* [to start] or *départ* [departure] but is intuitively not inceptive.

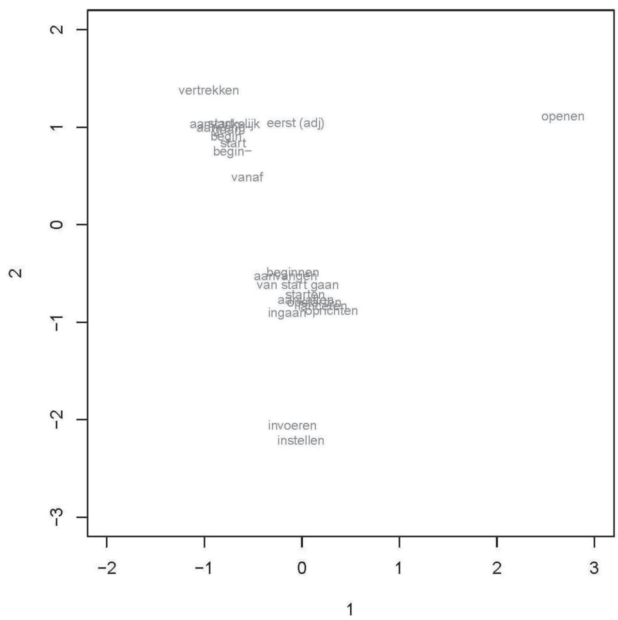


Figure 1: Semantic field of translated Dutch BEGINNEN

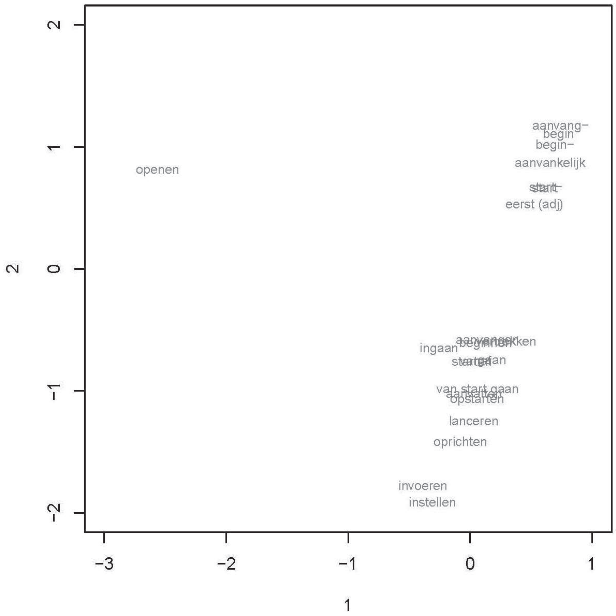


Figure 2: Semantic field of original Dutch BEGINNEN

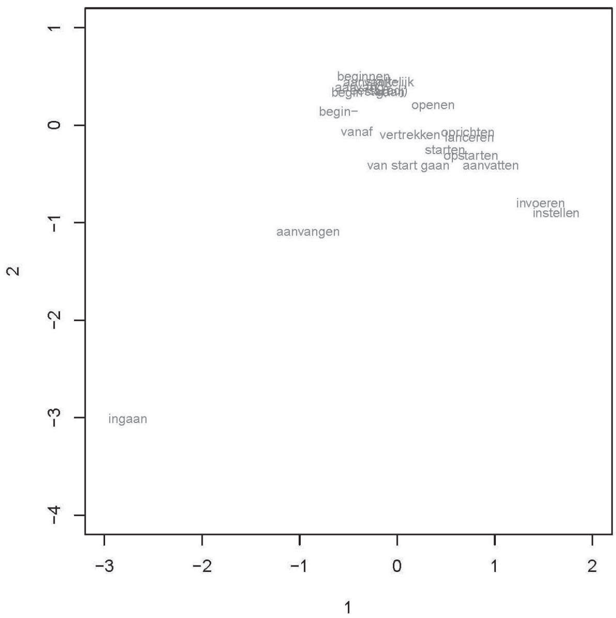


Figure 3: Semantic field of translated Dutch BEGINNEN – with anchoring

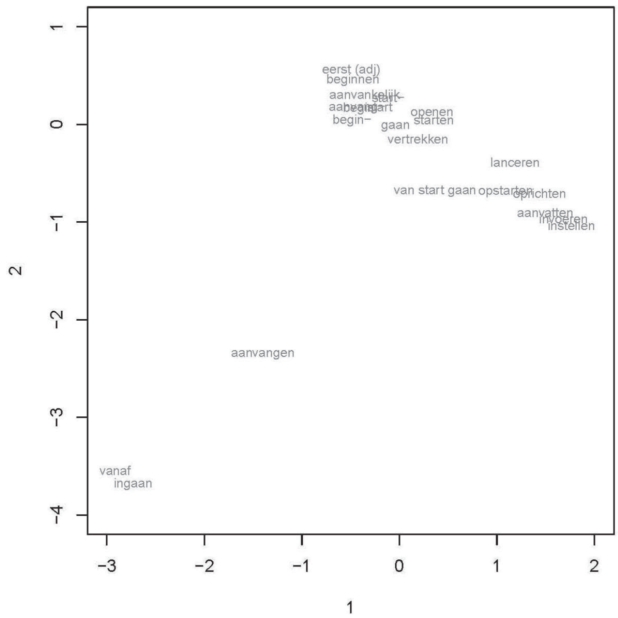


Figure 4: Semantic field of original Dutch BEGINNEN – with anchoring

5. Conclusion

In this article, we have made a first attempt to measure semantic differences between translated and non-translated language. We therefore have presented a quantitative bottom-up corpus-based method which enabled us to identify, to measure and to visualize semantic similarity between the elements in the field of Dutch inceptiveness. The method we applied is an extension to Helge Dyvik's Semantic Mirroring Technique, enabling the creation of both translated and non-translated semantic fields; it is translational and makes use of both a comparable and parallel corpus. We found that translational data are an interesting source for the bottom-up identification of a semantic field's structure and for the differentiation of prototypical meanings from peripheral ones. Moreover, the method enabled us to compare semantic fields in original and in translated language, while using the same data set. In a next step, we 'stabilised' the semantic field by 'anchoring' the positions of the 12 French lexemes. This approach revealed more meaning-differentiated semantic fields. The application of the back-and-forth technique to the semantic field of Dutch inceptiveness, did indeed show differences between the translated and the original field. The differences were especially laid bare by the anchored points technique. When we compared the two semantic fields, we observed a tendency towards a 'flatter' image in the translated field: the lexemes were plotted closer together, which indicated that the sense distinctions were less clearly present in the translated semantic field of inceptiveness. The created field is thus less fine-grained than when we use original language frequencies.

Our choice for the case of Dutch inceptiveness (BEGINNEN), might not have been the most appropriate one to test a method that is initially interested in conceptual meaning differentiation, for BEGINNEN is not really what one would call highly polysemous, making the sense distinctions less obvious, more subtle and thus harder to capture. Despite this somewhat unfortunate choice, the visualizations did show us detailed differences between and within the semantic fields and did enable a comparison between translated and original semantic fields. Meaning-differentiated semantic fields are thus very likely to corroborate the obtained results for BEGINNEN.

The proposed technique should be validated in several ways. A first validation can be done by implementing a different 'pivot-language', e.g. English (also available in the DPC). In this way, the semantic field of inceptiveness in both original and translated language can be created in the same way as proposed in this study, the only difference being the pivot language. A second step in the validation process can be made via the creation of semantic fields of inceptiveness in a different language (e.g. French), which enables us to make a cross-linguistic comparison of semantic fields and of the possible (language dependent or independent) influence of translation on semantic fields. This influence is not necessarily similar in different languages (for different languages have different attitudes towards translations) but we do expect to observe similar patterns. Finally, our translational method could be completed via aspects of distributional methods like the Behavioural Profile method (Divjak and Gries, 2006; Gries and Divjak, 2009). In computational linguistics, the assumption that words with a similar meaning tend to occur in similar contexts (Harris 1968) has led to the advent of distributional approaches like first and second order bag-of-words models (Manning & Schütze, 1999) and the behavioural profiles method (Divjak & Gries, 2006; Gries and Divjak, 2009). The combination of a context bound sense discrimination method with the translation based approach could provide us with a solid 'mixed' both distributional and translational instrument for the mapping of semantic networks.

Notes

1 This does not mean that the role of semantics itself in translation has not been addressed (e.g., Klaudy, 2010), but this kind of research is rarely corpus-based and barely ever involves with denotational issues.
2 Constrastive Linguistics have indeed used distributional methods for monolingual semantic differentiation (like the Behavioral Profile method developed by Divjak & Gries 2006; 2009). As the current study does not involve with distributional techniques, we will not elaborate further on this matter.
3 Note that we do not consider back-translations as translation-effect neutral. We do acknowledge that the technique rules out idiosyncratic translations and translations that are purely context-bound. Whether or not back-translations indeed rule out any translational effect, is exactly what emerges from our results.
4 de Clerck (1981); Reinsma (1993), Reinsma (1995), Boon and Geeraerts (2005), De Boer (2006), Van Dale (2010), Den Boon and Geeraerts (2011), Van Dale (2012).
5 Antonyms were excluded, though.
6 At every level of the back-and-forth translation technique, invalid alignments are eliminated from the data. Furthermore, if, in the translated sentence containing the lexeme under study, there is no translation equivalent (identifiable as such), the observation is not taken into account. In this way, we only take into account 'linguistically predictable translations' (Dyvik, 1998: 52).
7 The technical term in correpondence analysis is that the 22 Dutch lexemes are depicted as so-called "supplementary" or "illustrative" points.

References

Aijmer, K., and A.-M. Simon-Vandenbergen. (2004). "A model and a methodology for the study of pragmatic markers: the semantic field of expectation." *Journal of Pragmatics* 36 (10): 1781–1805. http://dx.doi.org/10.1016/j.pragma.2004.05.005.

Aijmer, K., and A.-M. Simon-Vandenbergen. (2006). *Pragmatic Markers in Contrast*. Amsterdam: Elsevier.

Altenberg, B., and S. Granger. (2002). "Recent trends in cross-linguistic lexical studies." In *Lexis in Contrast: Corpus-Based Approaches*, ed. B. Altenberg and S. Granger, 3–48. Amsterdam: John Benjamins. http://dx.doi.org/10.1075/scl.7.04alt.

Baker, M. (1993). "Corpus linguistics and translation studies. Implications and applications." In *Text and Technology. In Honour of John Sinclair*, ed. M. Baker, G. Francis, and E. Tognini-Bonelli, 233–245. Philadelphia, Amsterdam: John Benjamins. http://dx.doi.org/10.1075/z.64.15bak.

Baker, M. (2004). "A corpus-based view of similarity and difference in translation." *International Journal of Corpus Linguistics* 9 (2): 167–93. http://dx.doi.org/10.1075/ijcl.9.2.02bak.

Bernardini, S., and A. Ferraresi. (2011). "Practice, Description and Theory Come Together: Normalization or Interference in Italian Technical Translation?" *Meta* 56 (2): 226–46. http://dx.doi.org/10.7202/1006174ar.

Boon, T.d., and D. Geeraerts. (2005). "Van Dale: Groot Woordenboek der Nederlandse Taal: 3 Dl." In V. Dale (ed.) *Van Dale: Groot Woordenboek der Nederlandse Taal: 3 Dl*, 14th ed. Utrecht: Van Dale Lexicografie.

Dale, V. (ed.) (2010). *Van Dale thesaurus. Synoniemen en betekenisverwante woorden* (Eerste editie, eerste oplage ed.). Utrecht & Antwerpen: Van Dale Uitgevers.

Dale, V. (ed.) (2012). *Van Dale Onlinewoordenboek hedendaags Nederlands.*. Utrecht: Van Dale Uitgevers.

De Boer, W.T.e. (ed.) (2006). *Koenen Woordenboek Nederlands.* Utrecht & Antwerpen: Van Dale Uitgevers.

de Clerck, W. (ed.) (1981). *Nijhoffs Zuid Nederlands woordenboek. 's*-Gravenshage & Antwerpen: Martinus Nijhoff.

De Sutter, G., I. Delaere, and K. Plevoets. (2012). "Lexical lectometry in corpus-based translation studies. Combining profile-based correspondence analysis and logistic regression modeling." In M. Oakes and J. Meng (eds), *Quantitative Methods in Corpus-based Translation Studies. A practical guide to descriptive translation research*, 325–346. Amsterdam, Philadelphia: John Benjamins Publishing Company. http://dx.doi.org/10.1075/scl.51.13sut.

Delaere, I., G. De Sutter, and K. Plevoets. (2012). "Is translated language more standardized than non-translated language? Using profile-based correspondence analysis for measuring linguistic distances between language varieties. *Target*." *International Journal of Translation Studies* 24 (2): 203–24. http://dx.doi.org/10.1075/target.24.2.01del.

Den Boon, C.A., and D. Geeraerts eds. (2011). *Dikke Van Dale*. Utrecht: Van Dale Uitgevers.

Divjak, D., and S. Gries. (2006). "Ways of Trying in Russian. Clustering Behavioral Profiles." *Corpus Linguistics and Linguistic Theory* 2 (1): 23–60. http://dx.doi.org/10.1515/CLLT.2006.002.

Dyvik, H. (1998). "A translational basis for semantics." In S. Johansson and S. Oksefjell (eds), *Corpora and cross-linguistic research: theory, method, and case studies*, 51–86. Amsterdam: Rodopi.

Dyvik, H. (2004). "Translations as semantic mirrors from parallel corpus to wordnet." In *Advances in Corpus Linguistics*, ed. K. Aijmer and B. Altenberg, 311–326. Amsterdam, New York: Rodopi.

Greenacre, M. (2007). *Correspondence analysis in practice*. 2nd ed. Boca Raton: Chapman & Hall/CRC. http://dx.doi.org/10.1201/9781420011234.

Gries, S., and D. Divjak. (2009). "Behavioral Profiles. A Corpus-Based Approach to Cognitive Semantic Analysis." In V. Evans and S.S. Pourcel (eds), *New Directions in Cognitive Linguistics*, 57–75. Amsterdam, Philadelphia: John Benjamins. http://dx.doi.org/10.1075/hcp.24.07gri.

Halverson, S. (2003). "The Cognitive Basis of Translation Universals." *Target* 15 (2): 197–241. http://dx.doi.org/10.1075/target.15.2.02hal.

Halverson, S. (2010). "Cognitive translation studies: developments in theory and method." In G. Shreve and E. Angelone (eds), *Translation and cognition*, 349–369. Amsterdam: John Benjamins. http://dx.doi.org/10.1075/ata.xv.18hal.

Harris, Z.S. (1968). *Mathematical structures of language*. Wiley.

Ide, N., T. Erjavec, and D. Tufis. (2002). *Sense discrimination with parallel corpora.* Paper presented at the Proceedings of the SIGLEX/SENSEVAL workshop on Word sense disambiguation: recent successes and future directions, Philadelphia.

Ivir, V. (1983). "A Translation-based Model of Contrastive Analysis." *Jyväskylä Cross-Language Studies* 9:171–8.

Ivir, V. (1987). "Functionalism in Contrastive Analysis and Translation Studies." In *Functionalism in Linguistics*, ed. R. Dirven and V. Fried, 471–481. Amsterdam, Philadelphia: John Benjamins. http://dx.doi.org/10.1075/llsee.20.25ivi.

Klaudy, K. (2010). "Specification and Generalisation of Meaning in Translation." In B. Lewandowska-Tomasczyk & M. Thelen (Eds.), *Meaning in Translation* (Vol. 19, pp. 81–103). Frankfurt a.M: Peter Lang.

Kruger, H. (2012). "A corpus-based study of the mediation effect in translated and edited language." *Target* 24 (2): 355–88. http://dx.doi.org/10.1075/target.24.2.07kru.

Laviosa, S. (1998). "Core patterns of lexical use in a comparable corpus of English narrative prose." *Meta* 43 (4): 557–70. http://dx.doi.org/10.7202/003425ar.

Laviosa, S. (2002). *Corpus-based Translation Studies. Theory, Findings, Applications*. Amsterdam, New York: Rodopi.

Lebart, L., A. Salem, and L. Berry. (1998). *Exploring textual data*. Dordrecht: Kluwer Academic Publishers. http://dx.doi.org/10.1007/978-94-017-1525-6.

Lefever, E. (2012). *ParaSense: parallel corpora for word sense disambiguation*. Ghent: Ghent University.

Macken, L., O. De Clercq, and H. Paulussen. (2011). "Dutch Parallel Corpus: a Balanced Copyright-Cleared Parallel Corpus." *Meta* 56 (2): 374. http://dx.doi.org/10.7202/1006182ar.

Malmkjaer, K. (1997). "Punctuation in Hans Christian Andersen's stories and in their translations into English." In *Nonverbal Communication and Translation*, ed. F. Poyatos. Philadelphia: John Benjamins. http://dx.doi.org/10.1075/btl.17.13mal.

Manning, C.D., and H. Schütze. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.

Mauranen, A. (2000). "Strange strings in translated language. A study on corpora." In M. Olohan, (ed.), *Intercultural Faultlines: Research Models in Translation Studies I: Textual and Cognitive Aspects*, 119–141. Manchester: St Jerome.

Mortier, L., and L. Degand. (2009). "Adversative discourse markers in contrast: The need for a combined corpus approach." *International Journal of Corpus Linguistics* 14 (3): 338–66. http://dx.doi.org/10.1075/ijcl.14.3.03mor.

Noël, D. (2003). "Translations as evidence for semantics: an illustration." *Linguistics* 41 (4): 757–85. http://dx.doi.org/10.1515/ling.2003.024.

Olohan, M., and M. Baker. (2000). "Reporting that in Translated English: Evidence for Subconscious Processes of Explicitation?" *Across Languages and Cultures* 1 (2): 141–58. http://dx.doi.org/10.1556/Acr.1.2000.2.1.

Reinsma, R. (ed.) (1993). *Synoniemenwoordenboek*. Utrecht: Het Spectrum.

Reinsma, R. (ed.) (1995). *Prisma van de Synoniemen. Woorden met verwante betekenissen*. Utrecht: Het Spectrum.

Resnik, P., and D. Yarowsky. (1997). *A perspective on word sense disambiguation methods and their evaluation.* Paper presented at the ACL-SIGLEX Workshop Taggin Text with Lexical Semantics: Why, What and How? Washington D.C.

Resnik, P., and D. Yarowsky. (1999). "Distinguishing systems and distinguishing senses: New evaluation methods for word sense disambiguation." *Natural Language Engineering* 5 (2): 113–33. http://dx.doi.org/10.1017/S1351324999002211.

Simon-Vandenbergen, A.-M. (2013). "English adverbs of essence and their equivalents in Dutch and French." *Advances in Corpus-Based Contrastive Linguistics: Studies in Honour of Stig Johansson* 54:83–102. http://dx.doi.org/10.1075/scl.54.06sim.