

Nina Szymor

## Behavioral Profiling in Translation Studies

### *Abstract*

Behavioral Profiling offers an objective and precise corpus-based methodology that can be used for various cognitively-inspired lexical semantic analyses. The aim of this pilot study is to extend the application of Behavioral Profiling from lexical semantic studies within and across languages to the study of lexical semantics in translation. The focus is on analyzing instances of modality in translated and non-translated Polish texts to see whether translation affects the ways in which the words are used and their semantic networks are constructed. Data extracted from corpora (NKJP National Corpus of Polish and PELCRA Polish and English Language Corpora for Research and Applications) is annotated using Behavioral Profiling and analysed statistically to check whether the observed differences (if any) are significant. The results show variation in the use of modal words in translated and non-translated language, some of which can be explained with reference to the characteristics of human cognition.

### 1 Introduction

At the beginning of the 1990s, Mona Baker (1993) predicted that access to large corpora of translated and original texts will change the face of research in Translation Studies, particularly the study of principles of translational behaviour. She stressed the importance of developing appropriate tools that would enable this (Baker 1993: 235). This pilot study contributes to this quest by introducing a reliable, precise, and objective corpus-based methodology – Behavioral Profiling – already used in corpus-based lexical semantic studies within and across languages (Divjak 2004; Divjak/Gries 2006, 2009; Liu 2013; Deshors/Gries 2014; Divjak/Szymor/Socha-Michalik 2015).

To demonstrate how this methodology is used and what it can achieve, I will compare the use of two deontic modal verbs in texts translated into Polish with their use in texts originally written in Polish.

Firstly, some background information on corpus-based research in Translation Studies as well as the category of modality will be provided in this section. In Section 2, I will introduce the principles of the methodology and provide an overview of the study itself, explaining each of the steps taken in applying the methodology to translated and non-translated data. Finally, in Section 3, I will offer some comments and conclusions in relation to adopting this methodology in Translation Studies, as well as other

possibilities it offers.

### **1.1 Corpus-based Translation Studies**

Frawley (1984) put forward the idea of translated language being a language in its own right – a ‘third code’. This notion inspired the search for translation universals, i.e. “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). In other words, translation universals are features inherent to the translation process and translational behaviour – they are not influenced by the languages that the translator works with. Some of the hypothesized universals are:

- Simplification – the language or the message of the source text is simplified in the target text (Baker 1996: 176)
- Normalization – idiosyncratic features of the source text are transformed to conform to the conventions of the target language (Laviosa 2002: 54)
- Explication – information that is implicit in the source text is made explicit in the target text (Olohan 2001: 424)

Work on these universals started before the era of Corpus-Based Translation Studies (e.g. Blum-Kulka and Levenston 1983 look at simplification; Vinay and Darbelnet 1958/1977/1995 put forward the explication hypothesis; Vanderauwera 1985 studies normalization) but it was corpus-based research that transformed this intuitive and vague notion of translation universals into “clear, detailed operational research hypotheses” (Laviosa 2002: 75).

Some of the corpus-based studies dealing with translation universals still rely on frequencies and percentages only, without the use of confirmatory statistics. This is a major drawback – the results of these studies are based on a sample of the population of translated texts rather than the entire population so it is important to check their representativeness and validity (De Sutter/Van de Velde 2008: 2).

### **1.2 Modality**

It is generally agreed that modality relates to the attitudes and opinions of the speaker (Bybee/Perkins/Pagliuca 1994: 176). There are disputes in terms of the coverage of this category, as well as its internal structure – an agreement is yet to be reached on what belongs to the category of modality (e.g. should evidentiality, volition, desire, mood, etc. be included?), and how should the notions that undoubtedly belong to it be divided up into distinct categories (Nuyts 2006: 1-2). One traditional account distinguishes three modal categories (Nuyts 2006: 2-6):

- Dynamic modality – relates to the abilities and capacities (both inherent and imposed) of the participant of the clause, as well as his/her internal needs, e.g. That kid *can* sing like Frank Sinatra; I’ve unlocked the back door, so you *can* enter the house there.

- Deontic modality – relates to permission and obligation, including moral desirability, e.g. You *may* come in now; We *should* be grateful for what he has done for us.
- Epistemic modality – relates to the likelihood of the state of affairs being true, e.g. I *may* have put them on the table; they're not in the door (example from Bybee/Perkins/Pagliuca 1994: 180).

There are other classifications too – Van der Auwera and Plungian (1998) also recognise the existence of deontic and epistemic modality, but instead of dynamic modality they propose participant-inherent and participant-external categories, which closely relate to Nuyts' dynamic imposed and dynamic inherent modality. Van der Auwera and Plungian's deontic modality is just an extension (a case of a more 'specialised' use) of their participant-external modality. Bybee, Perkins and Pagliuca (1994) look at modality from a different perspective and apart from an epistemic sense, they propose an agent-oriented category, which covers states of affairs in which an internal or external condition compel the agent to complete the action in question, and a speaker-oriented category, in which the speaker grants an addressee permission (Bybee/Perkins/Pagliuca 1994: 177-180).

In this pilot study, I am looking at deontic modality as defined by Van der Auwera and Plungian (1998):

[deontic modality] identifies the enabling and compelling circumstances external to the participant as some person(s), often the speaker, and/or as some social or ethical norm(s) permitting or obliging the participant to engage in the state of affairs.

(Van der Auwera/Plungian 1998: 80)

More precisely, I am comparing the behaviour of two near-synonymous deontic modal verbs in legal texts originally written in Polish versus legal texts translated into Polish from English. The two verbs I look at are:

- *musieć* – 'to have to'; 'must'
- *powinien* – 'one should'

There are a number of other modal tools that express deontic modality in Polish – for a detailed overview I refer to Matulewska and Gortych (2009). In order to gain a good understanding of the behaviour and semantic network of these modals and to establish whether they differ from their translated counterparts a thorough study of all of them would have to be carried out. However, the main focus of this pilot study is to present how the Behavioral Profiling approach, already used in corpus-based cognitive semantic studies, can be successfully applied in Translation Studies, rather than to give a thorough account of the differences between translated and non-translated Polish texts in terms of use of deontic modals. For that reason, only two modal verbs will be the focus here.

## 2 Methodology

No two words mean the exact same thing and even the words that are considered to be synonymous look at the same situation from different perspectives (Divjak/Gries 2006: 24). The research methodology presented in this paper, i.e. Behavioral Profiling (Divjak 2004; Divjak/Gries 2006, 2009; Liu 2013; Deshors/Gries 2014; Divjak/Szymor/Socha-Michalik 2015), was developed out of the need for objective and reliable means to establish differences between the meanings of words. So far, it has been applied in various lexical semantic studies on near-synonymy (e.g. Divjak 2006 looks at verbs of 'intending' in Russian; Divjak and Gries 2006 look at nine near-synonymous Russian verbs of 'trying'; Gries and Otani 2010 look at the semantic field of 'size') and polysemy (e.g. Gries 2006 looks at the various senses of the word 'run'). This methodology can also be used in cross-linguistic studies, for example, Divjak and Gries (2009) look at the words that express 'begin' in English and Russian; they captured the behavioral profiles of each of the verbs and discovered that the prototypes in each language are based on different characteristics (e.g. the difference between 'begin' and 'start' is lexical, whereas the difference between their Russian 'counterparts' is aspectual or involve argument structure). In other words, there is no one-to-one mapping between the English and Russian verbs. This type of information can only be picked up by adequate methodology that can capture the complex nature of the phenomenon (Divjak/Gries 2009: 273). These findings can be used in, for example, the creation of bilingual and monolingual dictionaries, giving detailed information about (i) differences between synonyms in one language; (ii) usage-based (rather than intuition-based) functional translational equivalents.

Behavioral profiling is based on the assumption that distributional similarity correlates with functional and conceptual similarity. Another important assumption is that the choice of a near synonym is affected by its broader context, rather than just two contiguous words (Divjak/Gries 2006: 30). Finally, as mentioned in Section 1.1, simply relying on frequencies and percentages to make judgements about correlations in data may be misleading – certain relationships captured may occur by chance. In order to ensure that any regularities are significant, confirmatory statistics need to be applied. Following these three assumptions, the Behavioral Profiling method was developed: (i) a random sample of sentences is extracted from a corpus; (ii) all elements of each sentence are annotated for every clue possible; (iii) the annotated data is analysed statistically.

Divjak and Gries (2006) stress that additional research is needed to validate the findings provided by analysis of corpus data (Divjak/Gries 2006: 51-52). This could take the form of native speaker experiments, for example, sentence sorting tasks or gap filling experiments (Divjak/Gries 2008).

## 2.1 Sentence Extraction

Sentence extraction is the first step in applying Behavioral Profiling. Two corpora were used here for this purpose:

- National Corpus of Polish (NKJP, Narodowy Korpus Języka Polskiego 2008-2012) – a balanced, representative, morpho-syntactically annotated corpus of 239 million words. It consists of journalistic texts, belletristic literature, non-fiction, specialist periodicals and journals, other written texts, internet texts, and transcripts of conversations. The legal subcorpus, used in this study, consists of 7,161,072 words and contains (mainly) normative texts. This is the source of *non-translated data*.
- PELCRA English-Polish Parallel Corpora (s.a.) – a corpus of Polish translated from English. The JRC-Acquis subcorpus, used in this study, contains EU's *acquis communautaire* (28,571,342 target words). This is the source of *translated data*.

The difference in size between the corpora may be seen as a shortcoming. However, these are the only two corpora currently in existence for Polish that are comparable in terms of text type and timescale and can be used for the purpose of comparing translated and non-translated language. The smaller corpus is still large enough to give a representative sample of the use of the two verbs studied in non-translated legal Polish.

Sentences containing *musieć* 'have to', 'must' and *powinien* 'one should' were extracted from both corpora and manually validated. For authentic Polish, there were 614 sentences with *musieć* and 6,807 sentences with *powinien*. For translated Polish, there were 28,556 sentences with *musieć* and 24,952 sentences with *powinien*. Due to the difference in size between the two corpora, it makes more sense to convert these numbers into frequencies per million words; this way it will be easier to see the actual differences (if any) in frequencies between non-translated and translated language. Figure 1 shows that there is, in fact, a notable difference in the use of *musieć* – it occurs less frequently in original Polish (75.9 occurrences per million words) than translated Polish (999.4 occurrences per million words). *Powinien*, on the other hand, is used more in translated Polish (873.3 occurrences per million words) than in original (802.9 occurrences per million words), although the difference is not as staggering. As mentioned earlier, *powinien* and *musieć* are two of many modal tools used in legal Polish to express deonticity. Including others in this analysis would shed additional light on these frequencies, as well as other aspects of this study. However, giving a comprehensive overview of the modal situation in Polish is not the focus of this study, which aims at demonstrating how Behavioral Profiling can be successfully applied in Corpus-based Translation Studies. Thus, only two modals are analysed here, purely for reasons of demonstration.

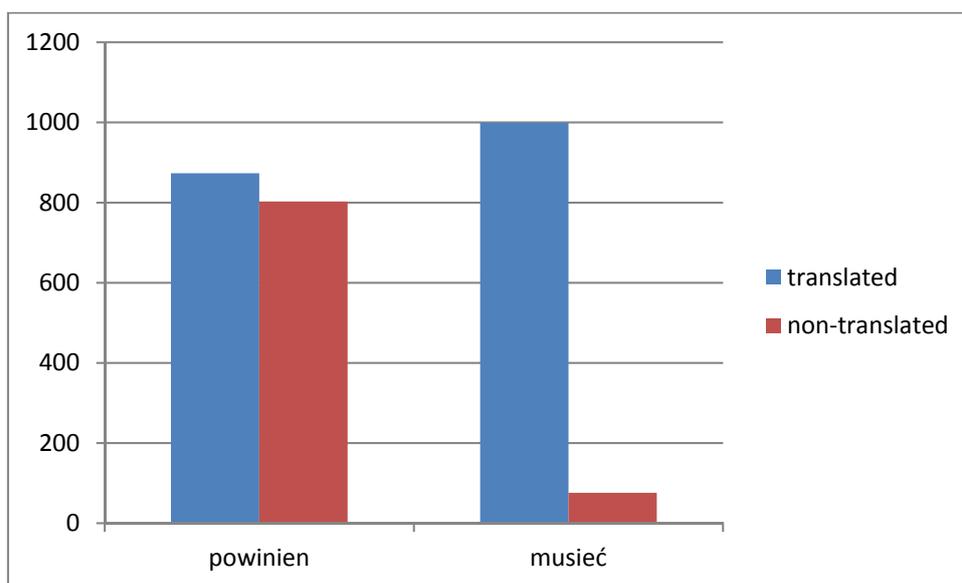


Fig. 1: Frequencies per million words

The sentences extracted from corpora were randomised, and a sample of 250 sentences for each of the verbs and each language type (i.e. non-translated and translated) were selected for analysis, resulting in a data base of one thousand sentences.

## 2.2 Annotation

The next step in the Behavioral Profiling approach is annotation. The annotation looks at every aspect of the sentence – first, the formal characteristics of the item in question are coded, e.g. if we are looking at the behaviour of a verb, we first look at its tense, aspect, and any other core characteristics. We then focus on the other elements of the sentence (mood, case of the subject slot, type of clause, etc.), including adverbs, particles, and connectors. Finally, semantic information on all elements of the sentence is encoded (e.g. animate vs. non-animate and concrete vs. abstract subjects and objects; a semantic classification of the item in question, etc.). All sentences are manually coded for the same number of parameters, making this approach precise and explicit; introspection and intuitiveness are removed until the stage of interpretation of findings.

Due to the legal nature of the texts under scrutiny in this study, the sentences are lengthy and complex. For that reason, the focus was on the immediate clause in which *musieć* or *powinien* appeared. The clauses were tagged for the following variables (all of which, with the exception of the semantic category of the infinitive, are objective and do not require intuition):

- clause (type: main/subordinate)
- subject (type: grammatical/compound/block/pronoun/logical/implied; animacy: animate/inanimate; common/proper; abstract/concrete),
- modal (tense: past/present/future; mood: indicative/subjunctive),
- infinitive (aspect: perfective/imperfective; voice: active/passive; semantic category [as in Divjak/Gries 2006: 34]: physical/physical other/exchange/motion/motion other/speech/intellectual).

Figure 2 illustrates the format in which the annotations were made. This is now ready for statistical analysis.

	A	B	J	K	L	M	N	O	P	Q	R	S
1	Sentence PL	Sentence EN	clause_type	sbj_pres	noun_anim	noun_type2	noun_type1	modal_tense	modal_mood	inf_aspect	inf_voice	inf_cat
2	Ekspert musi by	THE EXPERT M	main	gram	animate	common	concrete	present	indicative	n/a	active	phys
3	(2) Doświadczen	(2) Experience	main	gram	inanimate	common	abstract	present	indicative	n/a	active	phys
4	urządzenie to m	2.7. "Special f	main	gram	inanimate	common	concrete	present	indicative	imperf	active	phys
5	3. Jeżeli podczas	3. If, during ch	main	gram	inanimate	proper	abstract	present	indicative	perf	active	phys_other
6	Opinię przyjmuj	2. The represer	sub	gram	inanimate	proper	abstract	present	indicative	perf	active	phys_other
7	c) W przypadku	In the case set	main	gram	inanimate	common	concrete	present	indicative	imperf	active	speech
8	bez uszczerbku	c)Whereas, with	main	gram	inanimate	proper	abstract	present	indicative	imperf	być_passive	speech
9	w związku z tym	(9) Whereas pa	main	gram	inanimate	common	abstract	present	indicative	perf	zostać_passi	phys_other
10	"silnik urządzeni	-hand-held en	main	gram	inanimate	common	concrete	present	indicative	imperf	być_passive	phys_other
11	b) rynki regulow	(b) the regulat	main	gram	inanimate	common	abstract	present	indicative	perf	active	phys_other
12	W przypadku spi	(iv) a list of st	main	gram	inanimate	common	abstract	present	indicative	imperf	być_passive	phys_other
13	W szczególności	(8) Assessment	sub	gram	inanimate	common	abstract	present	indicative	imperf	być_passive	phys_other
14	3. W odniesieniu	3. As regards t	main	gram	inanimate	common	concrete	present	indicative	imperf	active	phys
15	5. Alkohol musi	5. The alcohol	main	gram	inanimate	common	concrete	present	indicative	perf	być_passive	phys_other
16	Równość taka st	3.1 The EESC c	main	gram	inanimate	common	abstract	present	indicative	n/a	active	phys
17	Ochrona ta mus	(31) In order tc	main	gram	inanimate	common	abstract	present	indicative	perf	być_passive	phys_other
18	Każde takie prze	2. However, wł	main	gram	inanimate	common	abstract	present	indicative	perf	być_passive	phys_other
19	c) większość, wy	(c) a majority -	main	gram	inanimate	common	concrete	present	indicative	n/a	active	phys
20	2. Kontrole zaw	2. Checks on th	main	gram	inanimate	common	abstract	present	indicative	imperf	active	phys
21	1.3 stwierdza, że	1.3 notes that	sub	gram	inanimate	common	concrete	future	subjunctive	perf	active	phys_other

Fig. 2: Annotated sentences in a spreadsheet

### 2.3 Statistical Analysis

Once the sentences have been annotated, statistical analyses need to be carried out to check for any significant interactions. In order to ensure that the regularities observed in my data are representative of the population as a whole and have not simply occurred by chance, it is crucial to run certain analyses.

There are various types of statistics that can be applied, depending on the purpose of the study and the type of data that one works with. In this study, Pearson's chi-

squared test was carried out and standardized residuals were analysed to check whether there are any significant correlations between translated and non-translated language in terms of how *musieć* and *powinien* are used, taking into account all of the characteristics (variables) described in Section 2.2.

Information about other statistics relevant for linguistic studies can be found in, among others, Baayen (2008) and Gries (2009).

#### *Pearson's Chi-squared and Standardized Residuals*

Pearson's chi-squared is a test of statistical significance, which tells you how likely it is that the results of your analysis are a matter of chance, or whether they are representative of what really goes on in the language. The frequencies that actually occurred in the data set are compared with frequencies that we would expect to occur if the only factor influencing them was chance. If the actual frequencies are close to the expected frequencies, then it is likely they occurred as a result of chance. The bigger the difference between the two values, the more chance that there are other contributing factors influencing and what we observed is an actual relationship. This information is provided in the form of a p-value (probability value), with a cut-off point of 0.05 – if p-value is less than 0.05, then we assume a significant relationship; if it is more than 0.05, then we know it occurred by chance (McEneary 2001: 84-85). In this pilot study, I want to establish which of the relationships between the two modals and the variables mentioned in the previous section are significant. For example, is the correlation between *powinien* and imperfective infinitives a result of chance or is there a significant relationship between them, meaning that *powinien* actually occurs more/less with imperfective infinitives?

Once it has been established which relationships are significant (with the Bonferroni correction applied), we can look at the standardized residuals to establish what the nature of those relationships is. If the standardized residual is a positive value then we have a positive association between two variables; if it is negative, the correlation is a negative one. Only values larger than 2/-2 point to significant associations/dissociations. To use the example of *powinien* and imperfective infinitives once more, I want to establish whether the relationship between them is a negative or a positive one. In other words, does *powinien* in translated texts occur with imperfective aspect significantly more, or significantly less often than in non-translated texts?

For this study, two analyses were carried out. Non-translated *musieć* was compared with translated *musieć* to establish whether there are any differences in how the modal behaves in texts originally written in Polish as opposed to Polish texts translated from English. Its behaviour in original Polish (its behavioral profile) serves as a benchmark, i.e. a standard use of the modal, against which the translated counterpart is compared. The same was done for *powinien*.

### 3 Results

As mentioned in the previous section, the behavioral profiles of non-translated modals were compared with their behavioral profiles in translated texts. The aim was to find out whether writers and translators use these two words differently. We look at every aspect of use, since every clue about the sentences in which they occurred was tagged.

#### 3.1 Translated vs. Non-translated *musieć*

There are several differences in how *musieć* is used in translated and non-translated texts in terms of clause type, animacy, type of subject, tense, voice and aspect. The differences are summarised in Table 1 and described below.

There are significantly more occurrences of *musieć* in subordinate clauses in non-translated (std. residual: 2.7) than in translated texts. Also in non-translated texts, animate subjects are more common (std. residual: 2.9) than in translated ones. In translated texts, on the other hand, there are significantly more occurrences of *musieć* in future tense (std. residual: 3.2) as well as occurrences where the modal verb refers to proper (std. residual: 2.9), as opposed to common, subjects.

The aspectual difference seems to be the most interesting one – there are clearly opposite preferences in translated vs. non-translated texts. *Musieć* is followed by an imperfective infinitive significantly more often in non-translated texts (std. residual: 2.1), whereas perfective infinitives occur more in translated texts (std. residual: 2.7). This is a very interesting outcome, considering studies on modality-aspect interaction in Slavic languages – it has been suggested that there is less chance of finding perfective infinitives than imperfective ones with deontic modals in Polish (Divjak 2011), as well as other Slavic languages (Šmelev/Zaliznjak 2006). Although the non-translated data used in this study supports those findings, the translated data shows the opposite tendency. The question needs to be asked as to why this is the case.

	subordinate clause	subject type: proper noun	future tense	animate subject	infinitive: pf aspect	infinitive: impf aspect
Translated	-2.7	2.9	3.3	-2.9	2.6	-2.1
Non-Translated	2.7	-2.9	-3.3	2.9	-2.6	2.1
P-value	0	0	0	0	0	0

Table 1: Differences observed for *musieć*.

#### 3.2 Translated vs. Non-translated *powinien*

Similarly to *musieć*, comparison of translated and non-translated *powinien* shows some differences in terms of several variables, including subject type, aspect, and semantic properties of the infinitive. In translated texts, infinitives relating to a physical action of

the subject occur significantly less frequently than in non-translated texts. However, the most significant differences seem to be related to the type of noun that the modal refers to, and to the aspect of the infinitive that the modal precedes. The differences are summarised in Table 2 and described below.

*Powinien* occurs with concrete subjects significantly more in non-translated Polish (std. residual: 5.0) than in translated Polish. Abstract nouns, on the other hand, co-occur with non-translated *powinien* less (std. residual: -4.8) than with translated *powinien*.

In terms of aspectual differences, a tendency similar to the one seen with *musieć* occurs. Here, however, it relates to the passive construction in which the infinitive occurs rather than to the active infinitive. In Polish, passive voice can take one of the following two forms (Kaleta 1995: 304):

- (1) the verb *być* ‘to be’ (imperfective) plus an *imperfective/perfective* passive participle;
- or
- (2) the verb *zostać* ‘to become’ (perfective) plus a *perfective* passive participle

The analysis shows that in translated texts, *powinien* is followed by a *zostać* passive construction (i.e. the perfective one) significantly more than in non-translated texts, but is followed by *być* passive constructions (i.e. the imperfective/perfective ones) significantly less. This again points to a preference towards perfectives in translated texts.

	subject type: concrete noun	subject type: abstract noun	voice: <i>być</i> passive	voice: <i>zostać</i> passive
Translated	-5.2	4.7	-2.8	5.3
Non-Translated	5.2	-4.7	2.8	-5.3
P-value	0	0	0	0

Table 2: Differences observed for *powinien*.

### 3.3 Interpretation

This precise approach captured several distributional differences in the use of deontic modals between translated and non-translated language. The reasons behind these differences may be complex and establishing them is outside the scope of this study, however, some ideas are briefly presented below.

When we translate, a lexical representation of a concept (i.e. word) in a source text is linked, via conceptual memory, to the lexical representation of the same concept in a target language (De Groot 1992: 392). That is, words in two languages are linked to the same conceptual memory, where the meanings of these words are stored. When we see or hear a word in one language, its semantic network, stored in the conceptual memory, is activated (i.e. the word’s various senses and links between them, as well as

other specifications, including extralinguistic and metalinguistic information related to the word). This semantic network is then linked with a word in the other language, which shares (at least some of) its aspects. De Groot (1992) explains how various words can share elements of semantic networks: the English word *idea* and Dutch *idee* share some elements of the same semantic network, but not all. *Idea* also shares some elements of meaning with Dutch *inzicht*, which in turn shares elements with English *insight*. These semantically related pairs of words – *idea* and *insight*, and *idee* and *inzicht* – share a few elements within and across languages (De Groot 1992: 394).

Halverson (2003) points out that most semantic networks are asymmetrical – some elements are more central, or prominent, while others are more peripheral (Halverson 2003: 216). As a result, we can say that each semantic network has two key gravitational centres: the category prototype (the most central meaning) and the highest level schema (the most general meaning that will describe each member of the network). She posits that when a semantic network is activated by a lexical representation in one of the languages, the network's most central meaning will exert what she terms as 'gravitational pull', which will 'make us' choose a term in the other language, corresponding to that central meaning most closely (Halverson 2003: 218). Let us take the following sentence as an example:

The expert *must* be a national of a member state (extracted from PELCRA)

When the translator comes across the word *must*, its semantic network will be activated – we assume it consists of three senses: dynamic (i.e. abilities, capacities, internal needs), epistemic (i.e. likelihood of state of affairs being true), and deontic (i.e. permissions, obligations, moral desirability). The context of the above sentence suggests a deontic meaning, so the deontic sense of the word *must* will be linked to the lexical representations of the same concept of deonticity in Polish; this could be *musieć*, *powinien*, and other deontic verbs/expressions not considered here. Based on Halverson's gravitational pull hypothesis, the item corresponding to the most salient (i.e. the most central/prototypical) meaning of deonticity will 'overshadow' the other items that likewise express it, but which are more peripheral. This more peripheral lexical representation could share other links with *must* which the more central lexical representation lacks, but because it is not the prototypical sense in the semantic network, it will not be the first one to be 'picked' as a translation. In the above sentence, *musieć* was chosen by the translator. 'Statistical analysis showed that in non-translated Polish *musieć* occurs with abstract subjects more frequently, and *powinien* occurs with concrete subjects. Here, the subject is a concrete noun – a person – yet *musieć* was chosen instead of *powinien*. Is it because *musieć* has a more central 'position' in the semantic network of deonticity than *powinien*, and therefore exerted gravitational pull? This is likely as a recent study on modality in Polish (Divjak/Szymor/Socha-Michalik 2015) suggested that *powinien* occurs in deontic and dynamic contexts (expressing necessity) but also in epistemic contexts (expressing probability). *Musieć*, on the other hand, occurs in deontic and dynamic contexts, expressing necessity only. This perhaps suggests that *musieć* is a prototypical lexical representation of the semantic network of

deonticity, whereas *powinien* is pushed to its periphery, since it shares some elements of the semantic network of epistemicity. *Musieć* occurs in translated texts much more (999.463 occurrences per million words) than in authentic texts (85.741 occurrences per million words), i.e., it is chosen by translators notably more often than by authors of original Polish texts, perhaps confirming its central position in the semantic network of deonticity and the gravitational pull hypothesis. Further extensive research as well as considering the other modals in the semantic network of deonticity would be required to support or refute this potential explanation of (some) differences between translated and non-translated texts. If successful, this would support Halverson's claim that translation universals can be explained with reference to the characteristics of human cognition (Halverson 2003: 197-198).

Not all of the differences between non-translated and translated texts found in this study could be explained by gravitational pull. It therefore needs to be stressed that gravitational pull in translation can be overridden by other motivations (Halverson 2003: 220-221).

A particularly interesting characteristic of the translated texts, in comparison with non-translated texts, is the modals' preference for perfective infinitives. As mentioned in Section 3.1, studies on the modality-aspect interaction showed that there is *less* chance of finding perfective infinitives than imperfective ones with deontic modals in Polish (Divjak 2011), as well as other Slavic languages (Šmelev/Zaloznjak 2006). This is true in the non-translated data – *musieć* is significantly more likely to be followed by an imperfective infinitive in comparison with translated texts, and *powinien* is significantly more likely to be followed by an imperfective passive construction. It is the opposite in translated texts – here we have a higher chance of finding a perfective infinitive following *musieć* in comparison with non-translated texts, and a significantly higher chance to find a perfective passive construction following *powinien*. One possible explanation of this obvious preference for perfective aspect in translated texts (in comparison with non-translated texts) could be the characteristics of the category of Slavic aspect. Imperfective aspect is traditionally viewed as signalling that the speaker's perspective lies within the described situation, while perfective puts his/her perspective outside of the situation, viewing it as a whole (Dickey 2000: 36). Perhaps translators subconsciously choose perfective as opposed to imperfective infinitives more than authors of original texts because by transferring into another language something that has already been written, as opposed to writing it originally, places their perspective externally to the situations described. Legislators who draw up laws have an internal perspective and that is why we may be more likely to find imperfective infinitives in texts written originally in Polish. Again, further research is required to validate this hypothetical explanation. However, if this claim is upheld, it would once again support the existence of characteristic features of translational behaviour.

## 4 Conclusions

Thanks to the Behavioral Profiling approach, I was able to profile the behaviour of the deontic meanings of the near-synonymous *musieć* and *powinien* in texts written originally in Polish, and compare the results with how the two verbs behave in texts translated into Polish. The differences found relate to various aspects of the sentence, and would be unnoticed if a less precise corpus-based methodology was used.

There are several advantages of using the Behavioral Profiling approach. Firstly, it focuses on every characteristic of the sentence/clause, in which the item under investigation occurs. This way, every difference at the sentence level – even the smallest one – can be captured and checked for significance, as seen in Section 3. Secondly, the statistical analyses applied to the annotated corpus data means that only the statistically significant differences are picked out and focused on in the interpretation of the results. Finally, annotating every sentence for the same set of parameters prevents intuitive and subjective judgements, making the approach both precise and objective.

The approach can be successfully used for lexical semantic analyses in Translation Studies, particularly with the aim to compare translated and non-translated language, giving detailed, precise, and objective results. If we look at the other Polish modals – not only deontic but also epistemic and dynamic – we can create a detailed semantic map of modality in Polish, with behavioural profiles for each individual modal. This can then be compared with the use of these modals in translated texts – as has been done in this study but on a much smaller scale – contributing to the quest for establishing the features of translational behaviour and universals of translation. This pilot study already showed interesting results and some potential support for universals of translation.

These findings can also be used for other purposes. As mentioned in Section 2, behavioural profiles can be used for lexicographical purposes. If on top of the semantic map of modality in Polish, we create maps for other languages (like Divjak and Gries 2009 did for English and Russian ‘begin’ – see Section 2 for more details), then we would be able to create resources for translators and language learners with a previously unattainable level of granularity and precision.

## References

- Baayen, R. Harald (2008): *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press
- Baker, Mona (1993): “Corpus Linguistics and Translation Studies: Implications and Applications.” Mona Baker, Gill Francis, Elena Tognini-Bonelli (eds): *Text and Technology, in Honour of John Sinclair*. Amsterdam: Benjamins, 233-252
- Baker, Mona (1996): “Corpus-based Translation Studies: The Challenges that Lie Ahead.” Harold Somers (ed.): *Terminology, LSP and Translation: Studies in Language Engineering, in Honour of Juan C. Sager*. Amsterdam: Benjamins, 175-186

- Blum-Kulka, Shoshana; Eddie Levenston (1983): "Universals of Lexical Simplification." Claus Faerch, Gabriele Kasper (eds): *Strategies in Interlanguage Communication*. London/New York: Longman, 119-140
- Bybee, Joan L.; Revere D. Perkins, William Pagliuca (1994): *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press
- De Groot, Annette (1992): "Bilingual Lexical Representation: A Closer Look at Conceptual Representations." Ram Frost, Leonard Katz (eds): *Orthography, Phonology, Morphology, and Meaning*. Amsterdam: North Holland, 389-412
- Deshors, Sandra C.; Stefan Th. Gries (2014): "A Case for the Multifactorial Assessment of Learner Language: The Uses of *may* and *can* in French-English Interlanguage." Dylan Glynn, Justyna Robinson (eds): *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*. Amsterdam/Philadelphia: Benjamins, 179-204
- De Sutter, Gert; Marc Van de Velde (2008): "Do the Mechanisms that Govern Syntactic Choices Differ between Original and Translated Language? A Corpus-based Translation Study of PP Placement in Dutch and German." Richard Xiao, Lianzhen He, Ming Yue (eds): *Proceedings of the International Symposium on Using Corpora in Contrastive and Translation Studies (UCCTS 2008)*. Zhejiang University, Hangzhou, 1-38 – [http://www.lancaster.ac.uk/fass/projects/corpus/UCCTS2008Proceedings/papers/de\\_Sutter\\_and\\_de\\_Velde.pdf](http://www.lancaster.ac.uk/fass/projects/corpus/UCCTS2008Proceedings/papers/de_Sutter_and_de_Velde.pdf) (12.12.2013)
- Dickey, Stephen (2000): *Parameters of Slavic Aspect: A Cognitive Approach*. Stanford: CSLI Publications
- Divjak, Dagmar (2004): Degrees of Verb Integration. Conceptualizing and Categorizing Events in Russian. Unpublished PhD dissertation. Leuven: Katholieke Universiteit Leuven
- Divjak, Dagmar (2011): "Predicting Aspectual Choice in Modal Constructions: A Quest for the Holy Grail?" Marcin Grygiel, Laura Janda (eds.): *Slavic Linguistics in a Cognitive Framework*. Frankfurt a. M.: Lang, 67-86
- Divjak, Dagmar; Stefan Th. Gries (2006): "Ways of Trying in Russian. Clustering Behavioral Profiles." *Journal of Corpus Linguistics and Linguistic Theory* 2 [1]: 23-60

**trans-kom**

**ISSN 1867-4844**

**trans-kom** ist eine wissenschaftliche Zeitschrift für Translation und Fachkommunikation.

**trans-kom** veröffentlicht Forschungsergebnisse und wissenschaftliche Diskussionsbeiträge zu Themen des Übersetzens und Dolmetschens, der Fachkommunikation, der Technikkommunikation, der Fachsprachen, der Terminologie und verwandter Gebiete.

Beiträge können in deutscher, englischer, französischer oder spanischer Sprache eingereicht werden. Sie müssen nach den Publikationsrichtlinien der Zeitschrift gestaltet sein. Diese Richtlinien können von der **trans-kom**-Website heruntergeladen werden. Alle Beiträge werden vor der Veröffentlichung anonym begutachtet.

**trans-kom** wird ausschließlich im Internet publiziert: <http://www.trans-kom.eu>

Redaktion

Leona Van Vaerenbergh  
University of Antwerp  
Arts and Philosophy  
Applied Linguistics / Translation and Interpreting  
Schilderstraat 41  
B-2000 Antwerpen  
Belgien  
[Leona.VanVaerenbergh@uantwerpen.be](mailto:Leona.VanVaerenbergh@uantwerpen.be)

Klaus Schubert  
Universität Hildesheim  
Institut für Übersetzungswissenschaft  
und Fachkommunikation  
Universitätsplatz 1  
D-31141 Hildesheim  
Deutschland  
[klaus.schubert@uni-hildesheim.de](mailto:klaus.schubert@uni-hildesheim.de)

- Divjak, Dagmar; Stefan Th. Gries (2008): "Clusters in the Mind? Converging Evidence from Near Synonymy in Russian." *The Mental Lexicon* 3 [2]: 188-213
- Divjak, Dagmar; Stefan Th. Gries (2009): "Corpus-based Cognitive Semantics: A Contrastive Study of Phrasal Verbs in English and Russian." Barbara Lewandowska Tomaszczyk, Katarzyna Dziwirek (eds): *Studies in Cognitive Corpus Linguistics*. Frankfurt a. M.: Lang, 273-296
- Divjak, Dagmar; Nina Szymor; Anna Socha-Michalik (2015): "Towards a Usage-based Categorization of Polish Modals." *Corpora and Quantitative Data in Slavic Linguistics*, special issue of *Russian Linguistics. International Journal for the Study of Russian and other Slavic Languages* 39 [3]: 327-349
- Frawley, William (1984): "Prolegomenon to a Theory of Translation." William Frawley (ed.): *Translation: Literary, Linguistic and Philosophical Perspectives*. London/Toronto: Associated University Presses, 159-175
- Gries, Stefan Th. (2006): "Corpus-based Methods and Cognitive Semantics: The Many Senses of *to run*." Stefan Th. Gries, Anatol Stefanowitsch (eds): *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*. Berlin/New York: Mouton de Gruyter, 57-99
- Gries, Stefan Th. (2009): *Statistics for Linguistics with R: A Practical Introduction*. Berlin/New York: Mouton de Gruyter
- Gries, Stefan Th.; Naoki Otani (2010): "Behavioral Profiles: A Corpus-based Perspective on Synonymy and Antonymy." *ICAME Journal* 34: 121-150
- Halverson, Sandra (2003): "The Cognitive Basis of Translation Universals." *Target* 15 [2]: 197-241
- Kaleta, Zofia (1995): *Gramatyka języka polskiego dla cudzoziemców*. Kraków: Uniwersytet Jagielloński
- Laviosa, Sara (2002): *Corpus-based Translation Studies. Theory, Findings, Applications*. Amsterdam/New York: Rodopi
- Liu, Dilin (2013): "Salience and Construal in the Use of Synonymy: A Study of Two Sets of Near-synonymous Nouns." *Cognitive Linguistics* 24 [1]: 67-113
- Matulewska, Aleksandra; Karolina Gortych (2009): "Translacyjne problemy wyrażania modalności deontycznej w tekstach aktów normatywnych w języku polskim, angielskim i greckim." Anna Mróz, Adam Niewiadomski, Monika Pawelec (eds): *Prawo i język. Proceedings of Prawo i język 2008, Warszawa*. Warszawa: Międzywydziałowe Koło Naukowe Kultury Języka Prawnego i Prawniczego Uniwersytetu Warszawskiego *Lingua Iuris*, 65-78
- McEnery, Tony (2001): *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press
- NKJP National Corpus of Polish (2008-2012) – <http://www.nkjp.pl/> (20.11.2015)
- Nuyts, Jan (2006): "Modality: Overview and Linguistic Issues." William Frawley (ed.): *The Expression of Modality*. Berlin/New York: Mouton de Gruyter, 1-26
- Olohan, Maeve (2001): "Spelling Out the Optionals in Translation: A Corpus Study." *UCREL Technical Papers* 13: 423-432
- PELCRA English-Polish Parallel Corpora (s.a.) – <http://pelcra.pl/> (20.11.2015)
- Šmelev, Alexej; Anna Zaliznjak (2006): "Aspect, Modality and Closely Related Categories in Russian." Unpublished talk at The Inaugural Meeting of the Slavic Linguistic Society in Bloomington, Indiana, 8-10 September 2006
- Van der Auwera, Johan; Vladimir A. Plungian (1998): "Modality's Semantic Map." *Linguistic Typology* 2 [1]: 79-124
- Vanderauwera, Ria (1985): *Dutch Novels Translated into English. The Transformation of a "Minority" Literature*. Amsterdam: Rodopi

Vinay, Jean-Paul; Jean Darbelnet (1958): *Stylistique comparée du français et de l'anglais*. 2nd ed. 1977. Paris: Didier – translation by Juan C. Sager, M.-J. Hamel: Jean-Paul Vinay, Jean Darbelnet (1995): *Comparative Stylistics of French and English*. Amsterdam/Philadelphia: Benjamins

### *Author*

Nina Szymor is a PhD student at The University of Sheffield, UK. In her PhD thesis, she looks at the influence of the process of translation on the use and semantic structure of core Polish modals. She applies the usage-based Behavioral Profiling approach and explains observed differences between translated and non-translated texts in terms of cognitive mechanisms typical of language in general.

E-mail: [nsszymor1@sheffield.ac.uk](mailto:nsszymor1@sheffield.ac.uk)

Website: <https://sheffield.academia.edu/NinaSzymor>

## Neu bei Frank & Timme

### TRANSÜD. Arbeiten zur Theorie und Praxis des Übersetzens und Dolmetschens

Herausgegeben von  
Prof. Dr. Klaus-Dieter Baumann,  
Prof. Dr. Dr. h.c. Hartwig Kalverkämper,  
Prof. Dr. Klaus Schubert

Susanne Hagemann: **Einführung in das trans-  
lationswissenschaftliche Arbeiten.** Ein Lehr-  
und Übungsbuch. ISBN 978-3-7329-0125-8.

Xenia Wenzel: **Die Übersetzbarkeit philoso-  
phischer Diskurse.** Eine Übersetzungskritik an  
den beiden englischen Übersetzungen von Hei-  
deggers *Sein und Zeit*. ISBN 978-3-7329-0199-9.

Ralph Krüger: **The Interface between Scientific  
and Technical Translation Studies and Cogni-  
tive Linguistics.** ISBN 978-3-7329-0136-4.

### FFF: Forum für Fachsprachen-Forschung

Herausgegeben von  
Prof. Dr. Dr. h.c. Hartwig Kalverkämper

Raimund Drommel: **Sprachprofilung – Grund-  
lagen und Fallanalysen zur Forensischen  
Linguistik.** ISBN 978-3-7329-0158-6.

Peter Kastberg: **Technik der Kondensation  
und der Expansion in der Technik.**  
ISBN 978-3-7329-0221-7

### Ost-West-Express. Kultur und Übersetzung

Herausgegeben von  
Prof. Dr. Jekatherina Lebedewa und  
Prof. Dr. Gabriela Lehmann-Carli

Jekatherina Lebedewa (Hg.), unter Mitarbeit  
von Anja Holderbaum: **Tabu und Übersetzung.**  
ISBN 978-3-7329-0034-3

### TTT: Transkulturalität – Translation – Transfer

Herausgegeben von  
Prof. Dr. Dörte Andres, Dr. Martina Behr,  
Prof. Dr. Larisa Schippel,  
Dr. Cornelia Zwischenberger

Cornelia Zwischenberger/Martina Behr (eds.):  
**Interpreting Quality: A Look Around and  
Ahead.** ISBN 978-3-7329-0191-3.

Vorankündigung:  
Mehmet Tahir Öncü: **Basiswissen für Dol-  
metscher – Deutschland und die Türkei.**  
978-3-7329-0154-8.

