

## CHAPTER 5

# How Europeans GIVE

## A two-layered semantic typology based on two parallel corpora

Natalia Levshina

F.R.S.-FNRS – Université catholique de Louvain

This study investigates how ten European languages from the Germanic, Romance and Slavic groups divide the semantic space of giving. The study is based on exemplars of events of giving extracted from two very different parallel corpora: Bible translations and film subtitles. This probabilistic exemplar-based approach allows for integration of the postclassical categorization models and semantic typology. The analyses show that there was more cross-linguistic variation in verbalization of GIVE in the subtitles than in the Bible translations, and also more cross-linguistic variation at the level of specific verbs than at the level of more abstract constructions. Moreover, there were significant associations between the constructional and lexical ways of expressing GIVE in all languages in the sample. Finally, the way these ten languages cut the semantic space of GIVE conforms to the genetic relationships between them only at the level of specific verbs, but not at the more abstract and iconic constructional level of categorization.

### 1. Introduction

Semantic typology studies how languages ‘carve up’ different conceptual domains. The overwhelming majority of such studies share two common characteristics. First, they are carried out at the level of types; second, these types are words. Classical examples are linguistic and anthropological studies of kin terminology (e.g. Nerlove & Romney 1967) and basic colour terms (Berlin & Kay 1969). However, the achievements of postclassical theories of categorisation in the late 20th century (see Murphy 2002) suggest that the preference for types (words) over tokens (individual instances of a word in context) is not unproblematic, both for onomasiological and semasiological approaches to semantics. From the onomasiological perspective, natural language categories are known to have fuzzy boundaries. As

a result, the membership of exemplars in one or the other category is probabilistic. From the semasiological perspective, categories tend to have a prototypical structure, with a centre and a periphery. For example, the Dutch synonyms *vernietigen* and *vernietigen*, which both mean ‘destroy’, have nearly identical semantic ranges, but different prototypes: the former is more typical when one speaks about physical destruction (e.g. to destroy a building), whereas the latter is more typically applied to abstract events (e.g. to annul an agreement) (Geeraerts 1988). Due to such gradability of category membership, cross-linguistic comparisons of lexemes (types) appear to be too coarse-grained.

The traditional way of lexical comparison should therefore be replaced by probabilistic models. One such alternative is the recent exemplar-based approach (e.g. Walchli 2010; Walchli & Cysouw 2012), which represents linguistic categories as clouds of exemplars that are mapped onto the common probabilistic space. The cross-linguistically salient semantic distinctions emerge inductively, for example, as clusters and dimensions on Multidimensional Scaling maps. Exemplars are usually taken from parallel corpora, most commonly, translations of the Bible. However, the impact of this highly specific data source on results remains an open question.

The second problem is the above-mentioned preference for the lexical level of categorization in semantic typology (see, however, some counterexamples in Evans 2010). This means that more abstract patterns are rarely compared, although, in accordance with Construction Grammar (Goldberg 1995), they can also be regarded as form-meaning pairings. Are there differences in the way languages cut the space, with the help of constructions at varying levels of abstraction? Are languages more or less lexically or constructionally similar for a given domain? These questions deserve answers.

The present paper is a case study of verbs that denote events of giving, or, in other words, express the semantic category GIVE, in ten European languages: Germanic (Dutch, English, German, Swedish), Romance (French, Italian, Portuguese, Spanish) and Slavic (Bulgarian and Polish). The analyses are based on a parallel corpus of two text types: Bible translations and film subtitles. To the best of my knowledge, no systematic quantitative comparison of lexical and constructional expressions of GIVE in European languages has been carried out to date. While one exception is the study conducted by von Waldenfels (2012), which focused on semasiological extensions of GIVE verbs in three Slavic languages (Czech, Polish, Russian), studies comparing larger numbers of European languages seem to be lacking. The aim of this paper is to fill this gap, as well as to test the following hypotheses, which relate to the afore-mentioned problematic issues:

1. Hypothesis 1. One can expect more cross-linguistic variation in the way GIVE is expressed in the subtitles than in the Bible translations, due to obvious differences in translation practices. The goal when translating a sacred text is to convey the meaning as closely as possible, although the specific strategy of course depends on the translation function: Some translations aim at preserving the Otherness of the Divine Author, and therefore tend to be very literal; on the other hand, some translations, especially the ones created by American evangelical communities and by missionaries who work with indigenous languages, are usually less literal and more ‘domesticated’, or ‘naturalized’, so that they can exist as independent texts that can be studied without additional commentary (de Vries 2007). In contrast, subtitle creators and film translators generally have greater freedom of self-expression.<sup>1</sup> Subtitles also have spatial constraints, which may influence a translator’s choice of linguistic expressions.
2. Hypothesis 2. One can expect more cross-linguistic variation in the verb-specific categorization of GIVE than at the more abstract constructional level. The latter is represented in this study by the ditransitive construction, which is defined, very broadly, as the presence of the Agent (Giver), the Recipient and the Theme as the arguments of a (ditransitive) verb (see Section 2.2). The motivation for this hypothesis is as follows. First, intuitively, one can expect more specific expressions to be less iconic and more arbitrary than more abstract categories. The former should therefore be more prone to cross-linguistic variation than the latter. Second, the range of possible lexical alternatives for events in a language is much greater than the inventory of the main clause-level constructional templates. This should allow for more variation in the lexical expression of GIVE on purely statistical grounds.
3. Hypothesis 3. There will be significant associations between the constructional and verb-specific ways of expressing GIVE in all of the languages under consideration. There is some cross-linguistic evidence that GIVE verbs are the most common verbs in ditransitive constructions, especially in its subtype, the double-object construction (Malchukov et al. 2010). It has also been observed that most simple clause constructions have special relationships with one particular verb, which best reflect their meaning (Stefanowitsch & Gries 2003; Goldberg et al. 2004; Ellis & Ferreiro-Jr. 2009). For example, there is abundant evidence that *give* plays the central role in the use and acquisition of the English ditransitive construction. However, cross-linguistic investigations of this association have been scarce.

---

1. One example is Goblin’s parody translations of films into Russian, which have become their own genre.

4. Hypothesis 4. The considered languages will cut the probabilistic space of GIVE according to the genetic relationships between those languages, both at the verb-specific and the constructional level. For example, Germanic languages would be expected to be more similar to one another in encoding the same events of giving than the Romance and Slavic languages.

The remaining part of the paper is organized as follows: Section 2 defines the semantic category GIVE and offers a brief description of the main lexical and constructional strategies of encoding such events in different languages. Section 3 describes the data and methods. In Section 4, the afore-mentioned hypotheses are tested through a series of quantitative analyses. Finally, Section 5 offers a summary and a discussion of the findings, and charts a path for future research.

## 2. The category GIVE and linguistic strategies of expressing it

### 2.1 Lexical strategies of expressing GIVE

As has been shown in Newman (1996, 1998a), the category GIVE is simultaneously experientially basic and semantically complex. It is experientially basic because it is one of the most frequent activities of everyday human life; it is semantically complex because it involves three entities: the GIVER, the RECIPIENT and the THING being transferred. Moreover, in prototypical instances of GIVE, there is interaction both between the GIVER and the RECIPIENT and between the RECIPIENT and the THING: the change of control over the THING, passing from the GIVER to the RECIPIENT, physical motion of the THING from the GIVER, and a number of other parameters (Newman 1996). Examples of prototypical giving are given in (1):

- (1) a. *He gave her a diamond ring.*  
 b. *Could you pass me the salt?*

The parameters of GIVE events may vary, which results in a multitude of possible semantic extensions of the verbs that express GIVE. For example, some verbs may denote interpersonal communication, when information is construed as the object of transfer, e.g. *give advice to someone* in English or *mem-bari-tahu* ‘TRANS-give-know’ = ‘to inform’ in Malay (Newman 1996: 16). In some cases, a GIVE verb does not in fact denote any transfer, but rather a situation in which the agent affects another entity, as in (2):

- (2) a. *Give your carpets a thorough cleaning in the spring.*  
 b. *She gave him a passionate kiss.*

Other common extensions include causation and enablement, as in (3):

- (3) a. *This music gives me a headache.*  
 b. *Daj pospat!* [Russian]  
 Give sleep  
 ‘Let me sleep.’

These extensions, which illustrate the semasiological perspective on the semantics of giving, are only a few typical examples of the polysemy of GIVE verbs in different languages. Onomasiologically, the semantics of GIVE can be realized in very many different lexemes, and languages differ in the ways in which they cut this semantic domain. For example, many European languages have representative sets of special lexemes for commercial transfer, e.g. *sell*, *lend*, *pay*, whereas the Dyirbal language, spoken in a non-monetary society, lacks these verbs; it instead makes other distinctions, as exemplified by its special verb pertaining to kinship obligations, *gibin* ‘to provide food for relatives’ (Dixon 1973). Regardless, most languages seem to have one central lexeme or morpheme with the meaning ‘to give’. One striking exception is Amele, along with all other languages from the Gun family of Papua New Guinea. These languages lack such a lexeme (Roberts 1998). Instead, markers of tense, aspect and mood, as well as those of subject and object agreement, appear in the position where a verb is expected.

- (4) *Ceb it-ag-a* [Amele] (Roberts 1998: 1)  
 betelnut 1SG:IO-2SG:SUB-IMP  
 ‘Give me the betelnut!’

According to Roberts, GIVE is so basic that it does not have to be expressed overtly (Roberts 1998: 29–31). In Construction Grammar terms, this means that the construction, which contains the GIVER, THING and RECIPIENT, is fully equivalent with the category of GIVE in Amele. The next subsection considers the constructional options for the encoding of GIVE in more detail.

In addition to the level and quality of semantic detail, onomasiological variation may arise from different construals of the same event. An act of transfer can be conceptualized not only as giving, but also as taking or receiving. For instance, an example (5a) of an event of metaphorical transfer of information taken from the film *Avatar* (see Section 3) was expressed by *give* in the original English version, but was translated into Swedish with the help of *få* ‘get’ (5b):

- (5) a. *You’ve given me good usable intel.*  
 b. *Jag har fått användbar information.*  
 I have got usable information  
 ‘I’ve got usable information.’

## 2.2 Constructional strategies of expressing GIVE

The main construction used for encoding GIVE is the ditransitive construction, which is defined here, very broadly, following Malchukov, Haspelmath and Comrie (2010). It consists of a (ditransitive) verb, an Agent argument, a Recipient-like argument and a Theme argument. Malchukov et al. (2010) do not make the formal expression of the arguments more specific because there is substantial variation in the way that predicates and arguments are encoded. For example, some languages (e.g. Swahili) have a double-object construction, in which the RECIPIENT and THING are both encoded as the Patient of the transitive construction. In other languages (e.g. Russian) the RECIPIENT has oblique case marking. Some languages use both strategies. For example, in English, the choice of the construction depends on a number of contextual factors, such as word class, semantic class, length and givenness/newness of the RECIPIENT and THING (Bresnan et al. 2007). Numerous other examples are available in Haspelmath (2011).

Ditransitive constructions (in this broad sense) also exhibit substantial semantic variation. The core meaning of physical giving seems to be shared by most European ditransitive constructions. Some common extensions of this prototype are transfer of information (6a), directed ballistic motion (6b) and benefactive events (6c), as in English:

- (6) a. *Tell me the news.*  
 b. *He threw her the ball.*  
 c. *Cry me a river.*

Some extensions are more distant from the prototype. For example, the English double-object ditransitive also expresses negative or non-occurring transfer:

- (7) a. *It cost me a fortune.*  
 b. *They denied her a visa.*

Similarly, the Dutch ditransitive double-object construction is still used in the function of possession deprivation (see Colleman 2009):

- (8) *Dat land ontnam hem zijn staatsburgerschap.*  
 That country took away him his citizenship  
 'His citizenship in that country was revoked.'

Yet another extension is found in the Russian and German ditransitive constructions, which can express malefactive events with external possessors:

- (9) a. *On slomal mne ruku.*  
 he broke me arm  
 ‘He broke my arm.’  
 b. *Sie hat ihm sein Herz gebrochen.*  
 she has him his heart broken  
 ‘She broke his heart.’

The ditransitive construction, however, is not the only one that can express GIVE. The benefactive or applicative construction (10b) is probably the most functionally similar, due to the similarity between the roles of the RECIPIENT and BENEFICIARY.

- (10) a. *He brought her a delicious chocolate cake.*  
 b. *He brought a delicious chocolate cake for her.*

Another near-synonymous construction is the caused motion construction [SUBJ V OBJ OBL], with similar semantic roles for the RECIPIENT and the GOAL. In (11b), both roles seem to be blended, making it difficult to say whether the sentence is an instance of the prepositional dative construction, as in (11a), or of the caused motion construction, as in (11c):

- (11) a. *I gave the ball to John.*  
 b. *I kicked the ball to John.*  
 c. *I kicked the ball to the field.*

In addition to such subtle modifications of the roles, the RECIPIENT OR THING can be left unexpressed. This may happen in cases for which the RECIPIENT OR THING have low discourse prominence, when the emphasis is on the action itself (12a) or due to social taboos (12b; cf. Goldberg 2005). This creates new opportunities for onomasiological constructional variation:

- (12) a. *Jesus drove off everyone who was buying and selling in the temple.*  
 b. *Ona v konce koncov dala emu.* [Russian]  
 She finally gave him  
 ‘She finally consented to sex with him.’

As these examples demonstrate, substantial variation exists at both the lexical and constructional levels of expressing GIVE. In the remaining part of the paper, these two levels and the relationships between them will be investigated with the help of quantitative methods, which are introduced in the next section.

### 3. Data and methods

The data used in this case study were collected from two very different parallel corpora. The first group was multilingual Bible translations collected by Marburg University (Mayer & Cysouw in press). These translations were aligned verse by verse. Automatic word alignment was also carried out with the help of the GIZA++ alignment software (Och & Ney 2003).<sup>2</sup> The second source was my own parallel corpus of film subtitles, which is part of a larger multilingual ParTy corpus (PARallel corpora for TYpology) currently being developed by the author. The subtitles were collected from <[www.opensubtitles.org](http://www.opensubtitles.org)>, a large online repository of subtitles uploaded by users. For this case study, I used subtitles for the six films listed below, which represent different genres and different original languages (given in parentheses):

*Avatar* (English)  
*Das Leben der Anderen*, aka *The Lives of Others* (German)  
*El Laberinto del Fauno*, aka *Pan's Labyrinth* (Spanish)  
 *Inglourious Basterds* (English, German, French, Italian)  
*Le fabuleux destin d'Amélie Poulain*, aka *Amélie* (French)  
*The Tourist* (English, French, Italian, Russian, Spanish)

The subtitles were sentence-aligned with the help of a subtitle alignment package from Jörg Tiedemann's OPUS software,<sup>3</sup>

To find exemplars of GIVE, I used the list of verbs from the frame *Giving* on FrameNet,<sup>4</sup> as well as all frames that inherit from it: *Commerce\_pay*, *Commerce\_sell*, *Lending*, *Submitting\_documents*, *Supply* and *Surrendering\_possession*. This list contained 36 English verbs, such as *give*, *bequeath*, *yield*, *leave*, *fob*, *pass* and *volunteer*, in 144 forms. The resulting contexts were manually checked for spurious hits. The contexts that did not contain events of literal or metaphorical transfer were discarded. For example, the ambiguous verb *leave* was retained only in such contexts as (13a), not (13b).

- (13) a. *His rich uncle left him a fortune.*  
 b. *He left the room.*

---

2. I am very grateful to Thomas Mayer for his help with word alignment.

3. I am indebted to Jörg Tiedemann for providing his brand-new version of alignment software and detailed instructions on using it. The program is downloadable at <<https://bitbucket.org/tiedemann/uplug>>

4. See <<https://framenet.icsi.berkeley.edu/fndrupal/home>> (11 November 2013).

After having found 77 contexts in the subtitles corpus, I randomly selected an equal number of English contexts from the Bible. Next, translations of these 154 English contexts were found with the help of alignment information. The multilingual exemplars were then coded for two types of information: (a) the lexical expression, i.e. the verb of transfer, and (b) whether the event of giving was encoded with the help of the ditransitive construction or not. In the latter case, a binary variable with the values of ‘Yes’ and ‘No’ was created. The lexical choices were coded both in a complete form and in a binarized form, which distinguished between the canonical GIVE verbs (English *give*, German *geben*, Dutch *geven*, Swedish *ge*, French *donner*, Italian *dare*, Spanish *dar*, Portuguese *dar*, Bulgarian *davam*, Polish *dawać*), on the one hand, and all remaining verbs, on the other hand. Binarization was performed in two slightly different ways. In a more ‘relaxed’ coding approach, derived and phrasal verbs with the same root (such as *give away* and *übergeben* ‘hand over’) were also considered canonical. In a stricter approach, derived and phrasal verbs with the same roots were considered to be separate lexemes. This resulted in several spreadsheets, with 154 exemplars as rows and ten languages as columns. Missing contexts and erroneous translations were coded as ‘NA’.

Of course, this English-source approach represents a limitation of the present study. By using only the English verbs to extract the exemplars of GIVE, the author ran the risk of ignoring some important functions of GIVE verbs in the other languages, especially some idiomatic uses of GIVE, such as (14) in Dutch:

- (14) *Die opmerking gaf veel onrust.* (Janssen 1998:270)  
 This remark gave much commotion  
 ‘That remark caused great commotion.’

Another idiomatic use of GIVE verbs is exemplified by the German construction *es gibt* (Newman 1998b). The decision to begin with the list of English verbs was a pragmatic compromise in light of the fact that not all languages in the sample had lexicographic data comparable to the English-language FrameNet.

To test the hypotheses presented in Section 1, I employed a mix of classic and innovative quantitative methods, some of which are popular outside of linguistics. For instance, I used Cohen/Light’s *kappa* ( $\kappa$ ), a statistic commonly used in behavioural sciences to measure interrater agreement, to test Hypotheses 1 and 2. Normally, interrater agreement shows how different subjects rate the same set of stimuli. For the present study, the authors of each translation were treated as subjects who had to choose a lexical and constructional category for 154 exemplars of GIVE. For Hypothesis 3, I computed the *phi* coefficient ( $\phi$ ) as a measure of association between the constructional and lexical choices for each language. To test Hypothesis 4, and also to interpret the results of the previous tests, I used Multidimensional Scaling based on the Gower (Hamming) distances between the

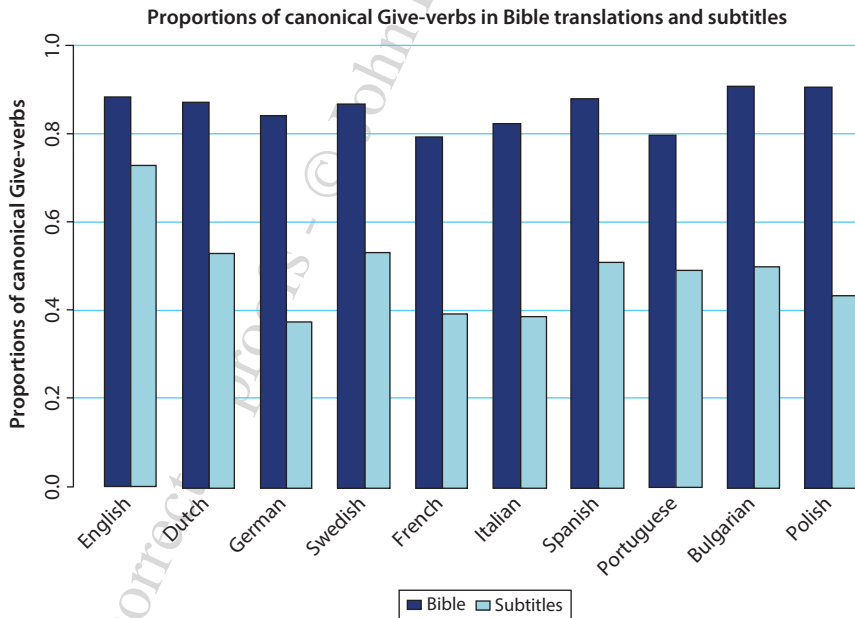
languages or exemplars. To define and compare the typological prototypes of GIVE in all languages and different language groups, I employed the Kriging algorithm; this algorithm was developed for geospatial statistics (see details below), but was recently applied by M. Cysouw to typology (Cysouw & Forker 2009).

For all analyses presented in the next section, I used R 3.0.1, a statistical environment and a programming language (R Core Team 2013), with the following add-on packages: *cluster* (Maechler et al. 2013), *fields* (Furrer et al. 2013), *psy* (Falissard 2012), *smacof* (de Leeuw & Mair 2009) and *vcd* (Meyer et al. 2012).

## 4. Quantitative analyses

### 4.1 Hypotheses 1 and 2

To get an initial sense of the data, I measured the proportions of the canonical GIVE verbs and ditransitive constructions in the two subcorpora for each language. Figure 1 shows the proportions of the canonical GIVE (the less strict definition, which treats the previous examples of *geben* and *übergeben* as members of the same group) for all languages in the two subcorpora. The proportions were

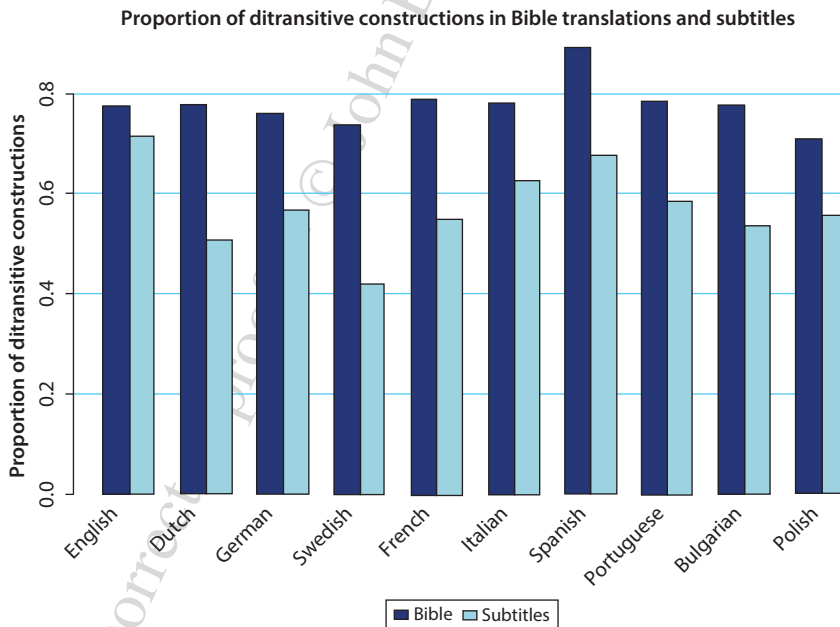


**Figure 1.** Proportions of the canonical GIVE verbs in the Bible translations and subtitles

computed because some contexts had a few missing translations, which would lead the raw frequencies to be misleading. The plot (Figure 1) shows a clear predominance of canonical verbs in the Bible translations. The average proportion of the canonical GIVE in the Bible translations is 0.857 (standard deviation  $sd = 0.042$ ), whereas the mean proportion in the subtitle data is almost two times smaller, at 0.487 ( $sd = 0.104$ ). The plot also shows greater variation in the frequencies of the canonical GIVE in the subtitles as compared to Bible translations, in terms of the afore-mentioned standard deviation values.

For constructions the situation is similar, although the difference between the two parallel corpora is somewhat less dramatic, as shown in Figure 2. The average proportion of the ditransitive constructions in the Bible translations is 0.778 ( $sd = 0.046$ ), whereas the mean proportion in the subtitles is 0.573 ( $sd = 0.084$ ). Again, the Bible translations exhibit more cross-linguistic similarity in the proportion of ditransitive expressions. There is substantial overlap between the ditransitive constructions and the canonical GIVE verbs, but it is not perfect (see Section 4.3).

However, the quantitative analysis of proportions alone does not suffice. It is also important to see how often the same situations are expressed in a similar way across the languages, and in which situations the languages tend to diverge. As mentioned in Section 3, since the sample consists of ten languages, I used



**Figure 2.** Proportions of ditransitive constructions in Bible translations and subtitles

Table 1. Fictional rater data illustrating *kappa* ( $\kappa$ )

Rater	Stimulus 1	Stimulus 2	Stimulus 3	Stimulus 4
A	X	X	Y	Y
B	Y	Y	X	X
C	X	X	X	Y

Light's *kappa*, a modified version of Cohen's *kappa* designed for multiple raters. This function is available through the R package *psy*. *Kappa* ( $\kappa$ ) is a coefficient of agreement between raters. It can be thought of as a measure of correlation for categorical variables. It ranges from  $-1$  (total disagreement) to  $1$  (total agreement). Consider the fictional case of three example raters, A, B and C, who were asked to assign four stimuli to two categories, X and Y, as shown in Table 1.

In this example, Raters A and B completely disagree, with different responses for all four stimuli ( $\kappa = -1$ ). Raters A and C agree in three cases out of four ( $\kappa = 0.5$ ), and Raters B and C agree in one case out of four ( $\kappa = -0.5$ ). Light's  $\kappa$  also computes such similarity between more than two raters, giving, roughly speaking, an average value of pairwise comparisons. For the example presented in Table 1, such a value would be equal to  $-0.33$ . When calculating Light's  $\kappa$ , missing values are usually omitted.

In the present study, the  $\kappa$  statistic is used as a measure of agreement between the languages in encoding GIVE; or rather, as a measure of agreement between the authors of translations. For the sake of comparability, the categorical variables were binary. For the lexical data, the values are 'Yes' if GIVE is expressed using a canonical GIVE verb (see the list in Section 3 and 'No' if the translator used a different lexeme. This design is similar to the one shown in Table 1, with X standing for a canonical GIVE verb and Y standing for anything else. This design allows one to compare how similarly these ten languages carve up the semantic space of giving (onomasiology), as well as how similar their canonical GIVE verbs are (semasiology). Both versions of binarization were used, the less strict method (e.g. *übergeben* = *geben* = 'Yes'; *verkaufen* 'sell' = 'No') and the more strict one (e.g. *geben* = 'Yes', *übergeben* = *verkaufen* = 'No'). The same approach was used to measure the constructional agreement between the translations (ditransitive = 'Yes', any other construction = 'No'). Table 2 presents the  $\kappa$  values for the two corpora and the two levels of analysis.

From the statistics in Table 2, one can infer the following. As for the text type variation, Hypothesis 1 is borne out, as the Bible translations agree more than the subtitles, both lexically and syntactically. Since the examined translations agree more in their constructional encoding than they do in their lexical ways of expressing GIVE, one can make a conclusion that Hypothesis 2 is borne out, as well.

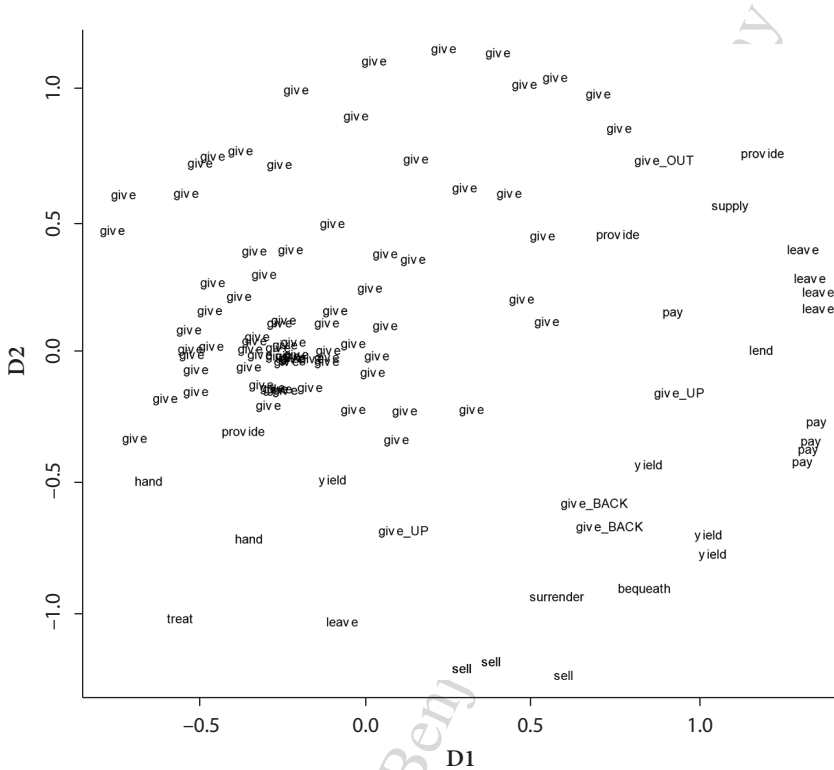
**Table 2.** Light's *kappa* statistics ( $\kappa$ ) for the two subcorpora and the two levels of categorization of GIVE events. The values in parentheses represent the stricter version of lexical comparison (e.g. *geben* = 'Yes', *übergeben* = 'No')

	Lexical	Constructional
Bible	0.57 (0.649)	0.721
Subtitles	0.339 (0.404)	0.530
Average	0.515 (0.588)	0.635

Notably, the stricter, whole word-match version of the lexical comparison yields higher agreement between the languages. One can hypothesize that the prefixal and phrasal verbs with GIVE are more prone to variation across these ten languages than the non-derived canonical GIVE verbs.

It is also interesting to explore the exemplars in categorization of which the languages agree most, as well as instances in which the translations most tend to disagree. In other words, what is the typological 'prototype' of the canonical GIVE verbs in these ten closely related languages? To find an answer, Gower distances – which are conceptually similar to well-known Hamming distances – were computed between all pairs of exemplars. The approach for calculating this value is very similar to the computation of  $\kappa$  statistics, but can only be applied to pairwise comparisons.<sup>5</sup> Comparisons were made between the actual lexemes, not the binarized values, and the derived distances reflected dissimilarities in the encodings of the same exemplar. Consider A and B, two exemplars of GIVE. Exemplar A contains *give* in the English version and *geben* 'give' in the German version. Exemplar B contains *give* in the English version and *verkaufen* 'sell' in the German version. The English verbs are identical, so the exemplars get one point added to their similarity score. The German translations are, however, different, so the two exemplars do not get any points added to their similarity score. The Gower similarity score ( $s$ ) is the sum similarity score divided by the total number of comparisons. In the above case of Exemplars A and B,  $s = 1/2 = 0.5$ . The higher the overlap of verbs in a pair of exemplars across all ten languages, the higher the similarity score of those exemplars. The dissimilarity score (distance) is simply a reverse similarity score ( $d = 1 - s$ ). All possible pairs of exemplars were compared for all ten languages (for 154 exemplars, a total of 11,781 pairs), and their similarity scores were obtained.

5. There is also a mathematical difference between the Gower similarity score and the  $\kappa$  statistic; namely, the former returns a simple proportion of agreement between two exemplars, while the latter also takes into account the probability of random agreement, which is important when raters are biased towards a specific category.



**Figure 3.** The probabilistic semantic space of GIVE based on the lexical information, with the English lexemes plotted on it

Partial similarity was taken into account by adding 0.5 in cases such as as *geben* and *übergeben* ‘hand over’.

The next step was to render the 11,781 distances into a low-dimensional space with the help of Multidimensional Scaling (MDS). A two-dimensional metric<sup>6</sup> solution was chosen as the most convenient one; this solution was also chosen because its representation of the observed distances between the exemplars was almost 95% correct. The scaling was done with the help of the *smacof* package in R, which usually returns low-stress solutions.

The result of MDS is a two-dimensional map shown in Figure 3. Following Wälchli and Cysouw (2012), I will call this map a probabilistic space, which serves as a tool for making cross-linguistic comparisons for the sample of GIVE exemplars. For convenience, the English labels are plotted onto the map. One can see

6. A non-metric solution turned out to be degenerate, see Borg and Groenen (1997), so the metric algorithm was used instead.

a dense cluster with *give* on the left, surrounded by less numerous other verbs. The cluster contains exemplars that denote not only physical giving (15), but also metaphorical transfer (16):

- (15) *And brought his head in a charger, and gave it to the damsel...*  
(King James Bible [KJB], Mark 6:28)
- (16) *I have given them thy word and the world hath hated them.* (KJB, John 17:14)

This dense clustering means that there is little cross-linguistic difference between literal and metaphorical instances of GIVE with regard to lexical categorization in this sample of languages. Some verbs have their own small clusters (e.g. *leave*, *pay*, *sell*). Below are two examples from the English translations:

- (17) *It's what pays for your science, compendo?* (*Avatar*)
- (18) *My aunt has left it to me.* (*Inglourious Basterds*)

The more compact the clusters, the more unanimous the languages are in encoding the events represented by these verbs. If exemplars of the same verb are located in different areas of the map, this often indicates polysemy. For example, the exemplars of *yield* in the bottom-right corner mean to 'surrender, to give up [to the control of the other]':

- (19) *Neither yield ye your members as instruments of unrighteousness unto sin...*  
(KJB, Romans 6:13)

The isolated exemplar of *yield* on the left (under the large cluster of *give*) represents the meaning 'produce':

- (20) *And some fell among thorns, and the thorns grew up, and choked it, and it yielded no fruit*  
(KJB, Mark 4:7)

This exemplar of *yield* is located near the large cluster of *give* because seven out of the ten languages use a canonical GIVE verb to express this meaning; the three exceptions are French with *rappporter* 'bring back', Italian with *fare* 'do, make' and English with *yield*.

Another example of polysemy is *leave*. Most exemplars with *leave* are located in the cluster on the right. The predominant sense is to entrust something to somebody or to stop interference, which only marginally resembles transfer, hence the position so far away from the *give* cluster:

- (21) *Leave the thinking to your superiors.* (*The Lives of Others*)

In the bottom left part of the map, however, one can find an exemplar of physical transfer:

(22) *He just left her an envelope. (The Tourist)*

This exemplar is closer to the *give* cluster than the other exemplars of *leave* because the German and Bulgarian translations encoded it with their canonical GIVE verbs *geben* and *davam*. In other words, this sense of *leave* is closer to the typological prototype of canonical GIVE verbs.

It is also possible to interpret the dimensions of the map in terms of the semantic characteristics of the main arguments (the GIVER, the RECIPIENT and the THING). Dimension 1 (D1, the horizontal axis) was found to correlate with the semantic class of the GIVER (adjusted  $R^2 = 0.08$ ,  $p_{F\text{-statistic}} = 0.001$ ) and, marginally, with that of the RECIPIENT (adjusted  $R^2 = 0.04$ ,  $p_{F\text{-statistic}} = 0.05$ ). Therefore, exemplars with human GIVERS and RECIPIENTS tend to be located on the left-hand side of the map. Exemplars with a non-human GIVER (23) and a non-human RECIPIENT (24) have a tendency to be located on the right-hand side of the map. This association is not particularly strong because non-human participants are relatively rare, especially with non-human RECIPIENTS.

(23) [a magic rose] *unable to bequeath its gift to anyone... (Pan's Labyrinth)*

(24) *...without trying to give his stifled life // the breath of air she had given to so many others... (Amelie)*

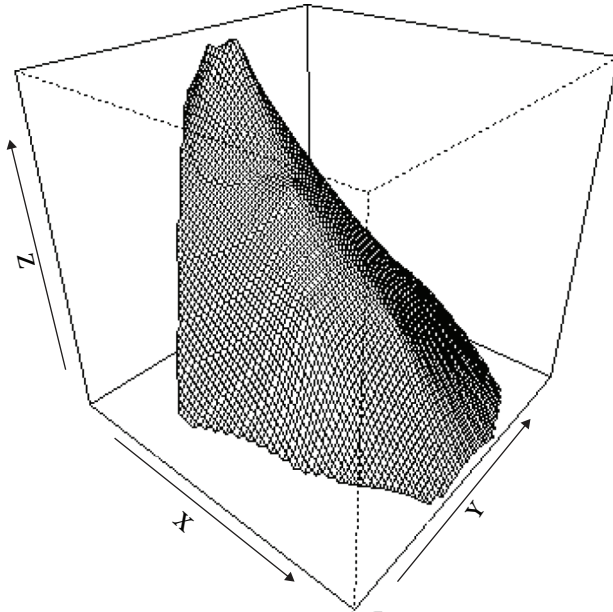
The second dimension (D2, the vertical axis) was found to correlate significantly with the semantic classes of the THING (adjusted  $R^2 = 0.14$ ,  $p_{F\text{-statistic}} < 0.0001$ ). Material THINGS (25) and less frequent human objects of transfer (26) tend to be located in the lower half of Figure 3, while abstract THINGS tend to be concentrated in the upper half of the map (27).

(25) *He just left her an envelope. (The Tourist)*

(26) *I'm ready to give you Pierce. (The Tourist)*

(27) *I'm getting all emotional, might just give you a big wet kiss. (Avatar)*

Up to this point, only English lexemes have been discussed. Let us now explore the other languages. The aggregate lexical choices in the ten languages can be represented with the help of Kriging, a geospatial technique that allows users to extrapolate and interpolate numeric values in space. This function is available in the R package *fields*. The result is shown in Figure 4. The axes  $x$  and  $y$  are the dimensions D1 and D2 of the MDS solution, respectively. The horizontal plane is equal to the MDS map that was just discussed. The new element here is the vertical dimension, which represents the density of exemplars with the canonical GIVE verbs in a specific region of the map. The peak is the where the canonical verbs



**Figure 4.** A perspectival plot of density of canonical GIVE verbs in the semantic space based on Kriging ( $\lambda = 0.01$ )

occur the most, and the low region is where none of the languages use canonical GIVE verbs.

The plot in Figure 4, which is based on multilingual information, shows basically the same pattern as that seen in the previous plot with the English labels (Figure 3), meaning that the languages tend to agree when they categorize GIVE. Since the peak is located on the left-hand side of the semantic map, which represents Dimension 1, and more in the middle with regard to Dimension 2, Figure 4 suggests that, on average, canonical GIVE verbs in the ten languages do not have any preferences with regard to the nature of the THING, but do prefer the human GIVER and RECIPIENT.

After interpreting the map, it is necessary to explore if there is any difference between the distributions of the exemplars from the Bible translations and those from the subtitles in the semantic space. Figure 5 shows the same map as Figure 3; however, in Figure 5 the exemplars from the Bible are represented as crosses, and those from the subtitles are represented as circles. Clearly, the overwhelming majority of the Bible exemplars contain canonical verbs of giving, while the exemplars from the subtitles exhibit much greater variation. This finding can be regarded as further evidence supporting Hypothesis 1.

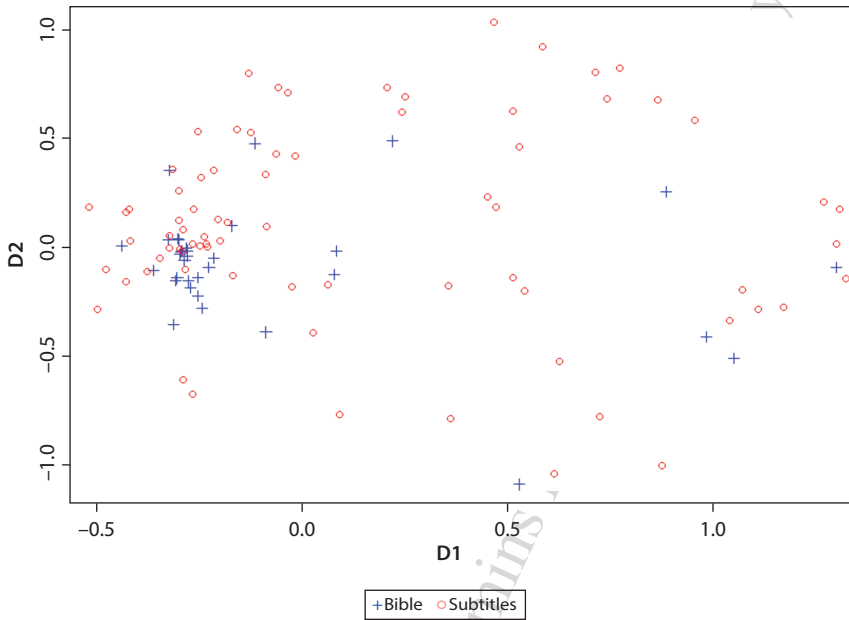
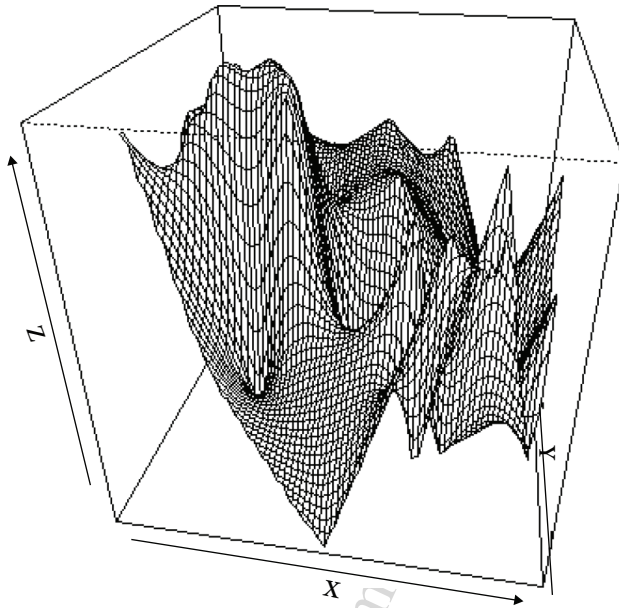


Figure 5. Comparison of distributions of exemplars from Bible translations and subtitles on the semantic map

#### 4.2 Hypothesis 3

Hypothesis 3 postulates that one should expect a strong association between the lexical and constructional choices in the sample languages; specifically, the canonical GIVE verbs should be closely associated with the ditransitive construction. Figure 6 shows the frequencies of the ditransitive construction in the ten languages, plotted onto the same semantic space represented by Figures 3, 4 and 5. The peaks correspond with the contexts where the ditransitive construction is selected in all languages, and the valleys correspond to the exemplars of other constructions. This irregular pattern vividly demonstrates that there is no one-to-one correspondence between the lexical and constructional choices. At the same time, one can also see that the greatest peak corresponds with the typological prototype of canonical GIVE verbs from Figure 4.

The smaller peaks on the right correspond with the clusters of verbs of leaving and yielding. The plot also indicates some association between the canonical GIVE verbs and the ditransitive construction.



**Figure 6.** The frequencies of the ditransitive construction in the ten languages, mapped onto the lexical space of GIVE. The peaks correspond with high frequencies of the ditransitive construction, the low regions to low frequencies of the ditransitive construction

To test this association between the canonical GIVE verbs and the ditransitive construction more formally, I computed *phi* coefficients ( $\phi$ ),<sup>7</sup> which measure the association between two categorical variables, with values ranging from 0 (no association) to 1 (perfect association). The results for each language are shown in Table 3 (for both types of texts). All associations were statistically significant ( $p < 0.05$ ) according to the Chi-squared test.

**Table 3.**  $\phi$  coefficients for associations between the lexical and constructional levels for the ten languages. The values in parentheses represent the stricter version of lexical comparison (e.g. *geben* = ‘Yes’, *übergeben* = ‘No’)

Language	$\phi$	Language	$\phi$
English	0.353 (0.453)	Italian	0.397 (0.397)
Dutch	0.591 (0.597)	Spanish	0.319 (0.319)
German	0.241 (0.277)	Portuguese	0.441 (0.441)
Swedish	0.557 (0.605)	Bulgarian	0.342 (0.469)
French	0.515 (0.504)	Polish	0.259 (0.407)

7. In this case, they are equal to Cramer’s *V* statistic.

All  $\phi$  coefficients in Table 3 have weak or moderate values,<sup>8</sup> and the Romance languages display less variation in their scores than do the Germanic languages. The stricter version of lexical comparison (e.g. *geben* = ‘Yes’, *übergerben* = ‘No’), in parentheses, yields higher coefficients, probably because prefixal derivations and phrasal verbs tend to appear in other constructions. This distinction does not play an important role in the Romance languages. This proves that, although there is a significant association between the constructional and lexical levels of categorization, it is not very strong.

### 4.3 Hypothesis 4

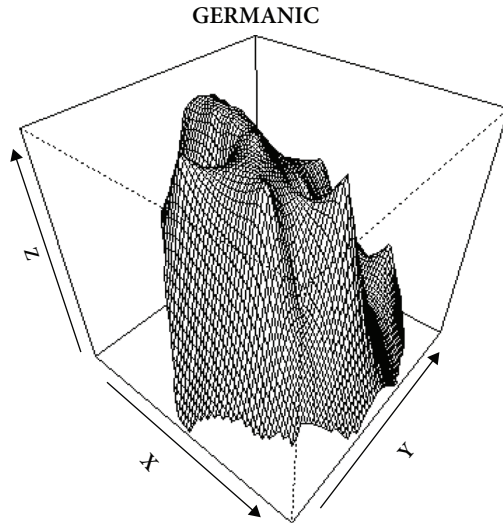
In this section the focus moves on to Hypothesis 4, and to exploring the genetic similarities between the sample languages in encoding GIVE. The first goal is to identify the prototypes of canonical GIVE verbs in the three language groups. Figures 7a, 7b and 7c show the peaks and valleys of the canonical GIVE verbs for the Germanic, Romance and Slavic groups represented in the study sample (the less strict version of lexical comparison was used). Some clear differences are evident: the canonical GIVE verbs in the four Romance languages tend to be located on the far left of the semantic space, while the Germanic verbs appear less restricted in their location.

Figures 8a to 8c are based on the stricter version of lexical comparison, which yields some changes for the Germanic and Slavic language groups. However, the previous conclusions still hold: the Germanic GIVE verbs are again closer to the centre, while the Romance cognates are more to the left, spread along the second dimension (the  $y$ -axis). The results for Bulgarian and Polish seem to be very affected by removing the prefixal verbs derived from the canonical *davam* and *dawać*; without the prefixal derivatives; the semantic range of their canonical verbs is quite modest. This is because many verbs of giving – for example, verbs of selling in these and other Slavic languages – are derived from the canonical verbs with the help of prefixes.

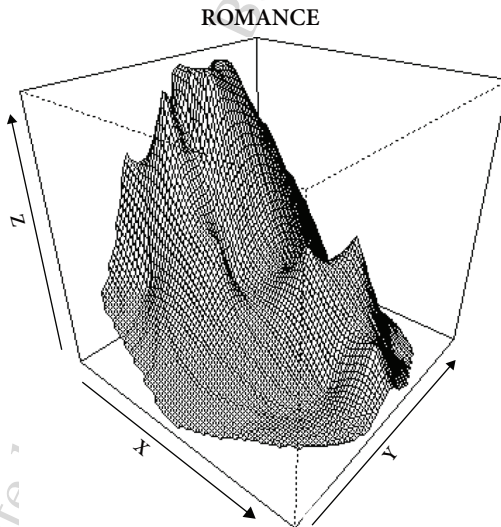
The next step in the data analysis is to create a common space for the languages and see how they cluster, which gives information regarding the behaviour of the individual languages. To do this, two lexical models and two constructional models are created for each subcorpus. The distances between the languages are

---

8. A separate test of the Bible and subtitle data sets shows that, on average, the subtitles yield somewhat higher coefficients of association, while the Bible translations yield lower scores. One explanation is that this is due to the high predominance of canonical GIVE verbs and ditransitive constructions in the Bible translations, which results in very low frequencies of non-canonical and non-ditransitive instances in this data set.



**Figure 7a.** Prototypes of canonical GIVE verbs from the three different language groups, using the less strict version of lexical comparison (e.g. *geben* = *übergeben* = ‘Yes’). This figure represents the Germanic languages



**Figure 7b.** Prototypes of canonical GIVE verbs from the three different language groups, using the less strict version of lexical comparison (e.g. *geben* = *übergeben* = ‘Yes’). This figure represents the Romance languages

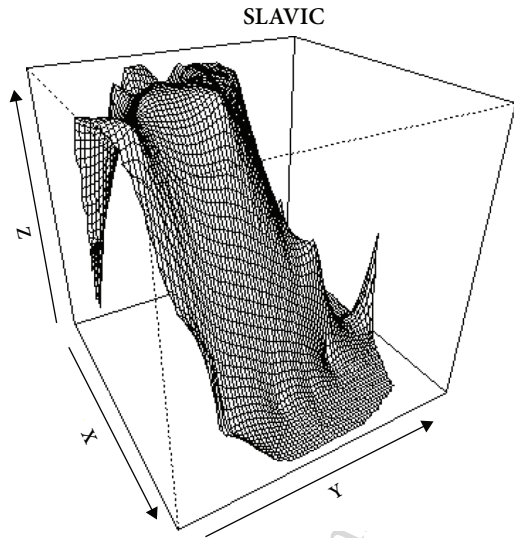


Figure 7c. Prototypes of canonical GIVE verbs from the three different language groups, using the less strict version of lexical comparison (e.g. *geben* = *übergeben* = 'Yes'). This figure represents the Slavic languages

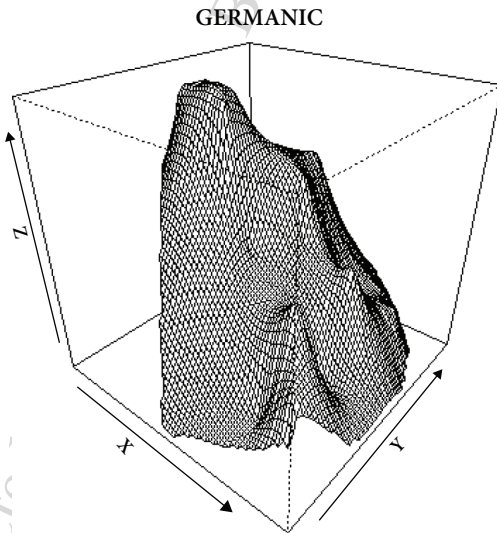
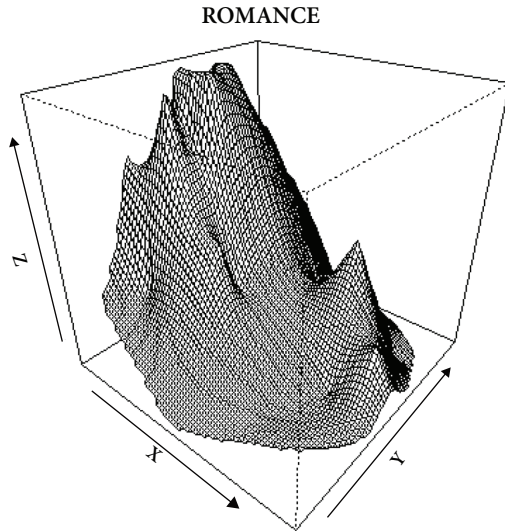
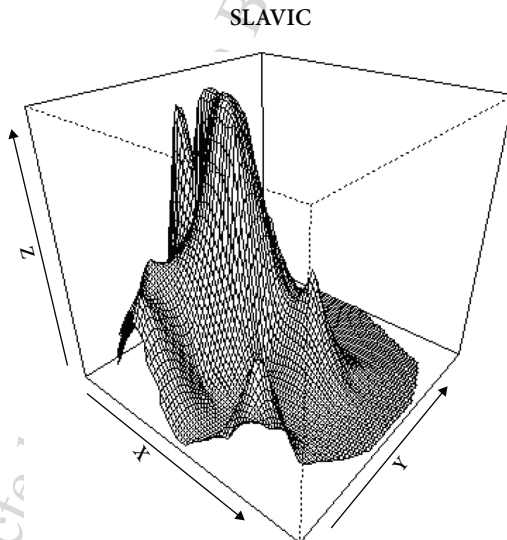


Figure 8a. Prototypes of canonical GIVE verbs from the three different language groups, using the stricter version of lexical comparison (e.g. *geben* = 'Yes', *übergeben* = 'No'). This figure represents the Germanic languages



**Figure 8b.** Prototypes of canonical GIVE verbs from the three different language groups, using the stricter version of lexical comparison (e.g. *geben* = 'Yes', *übergeben* = 'No'). This figure represents the Romance languages

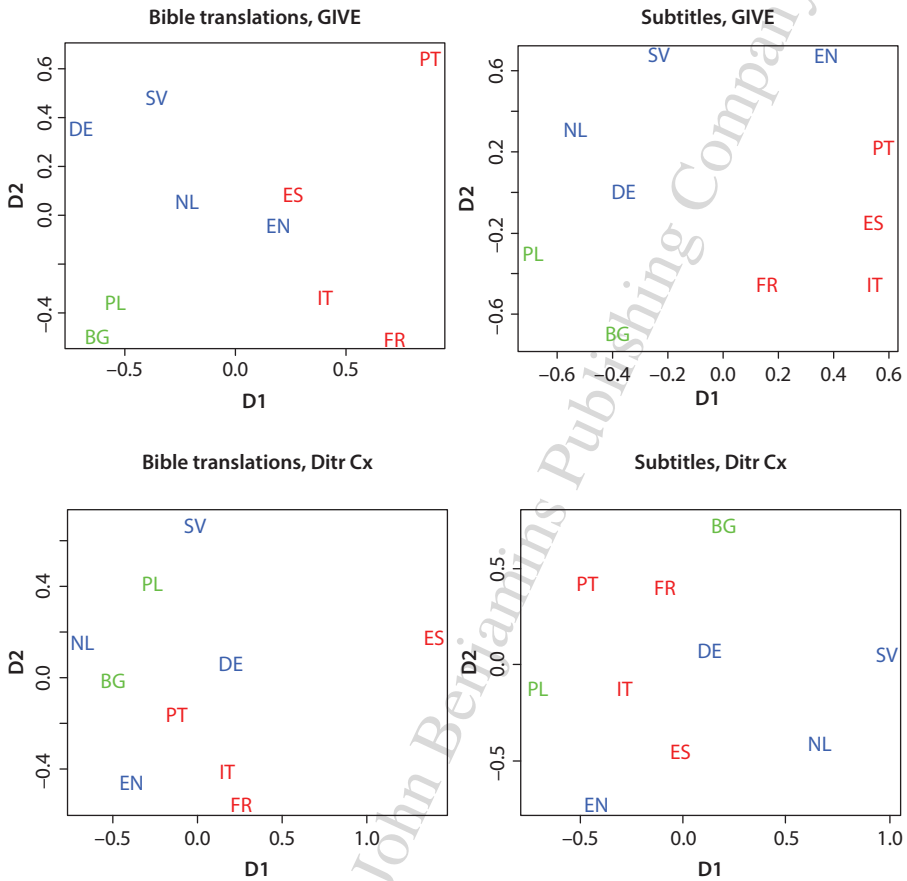


**Figure 8c.** Prototypes of canonical GIVE verbs from the three different language groups, using the stricter version of lexical comparison (e.g. *geben* = 'Yes', *übergeben* = 'No'). This figure represents the Slavic languages

defined by transposing the lexical and constructional matrices and applying the same approach as in the exemplar-based space. However, for this latter analysis, pairs of languages are compared, as opposed to exemplars. The asymmetric approach is used, meaning that one point is added to the similarity score between two languages if their values for a specific exemplar are both ‘Yes’ (the canonical GIVE verb if comparing lexemes, and the ditransitive construction if comparing constructions). If both languages have values ‘No’, nothing is added to their similarity score. These similarity scores are then transformed into distances, and MDS maps are built, as described in Section 4.1. The resulting plots are shown in Figure 9. The principle of interpretation is simple: the closer the languages, the more similar they are in lexically or constructionally encoding GIVE. The two plots on the left-hand side represent the Bible translations, and the two on the right represent the subtitles. The two upper plots show the similarities between the ten languages based on their lexical categorization, using the less strict approach, while the two lower plots are based on their constructional encoding. The upper plots compare the canonical GIVE verbs in the ten languages, while the labels on the lower plots correspond to the language-specific ditransitive constructions. The quality of the solutions is acceptable: the stress values range from 0.012 (Bible, lexical) to 0.053 (subtitles, lexical).

It is evident that the lexical solutions in Figure 9 yield results that are can be more easily interpreted along genetic lines: the Germanic (Dutch, English, German, Swedish), Romance (French, Italian, Spanish, Portuguese) and Slavic languages (Bulgarian, Polish) tend to be closer to one another than to the members of the other linguistic groups, although this tendency is not perfect. The most clear-cut solution seems to be that presented in the top-right plot (Figure 9: Subtitles, GIVE), while the constructional patterns (Figure 9: Bible Translations, Ditr Cx; Subtitles, Ditr Cx) are much less interpretable.

Although constructional categorization generally displays greater cross-linguistic similarity than lexical choices (as was shown in Section 4.1), constructional categorization is less associated with the genetic relationships between the languages. The constructional level, as defined in this paper, is more abstract and iconic than the lexical level, and is therefore less influenced by the formal and genetic differences between the ten languages. For more specific descriptions of constructions, such as argument marking or word order, a more genetically interpretable clustering solution might be expected.



**Figure 9.** Multidimensional Scaling of interlingual similarities in encoding GIVE in the Bible (left) and in the subtitles (right), based on lexical encoding (top) and constructional choice (bottom). ‘BG’: Bulgarian; ‘DE’: German; ‘EN’: English; ‘ES’: Spanish; ‘FR’: French; ‘IT’: Italian; ‘NL’: Dutch; ‘PL’: Polish; ‘PT’: Portuguese; ‘SV’: Swedish

## 5. Summary and outlook

This quantitative corpus-driven investigation of lexical and constructional expression of GIVE in two different parallel corpora has yielded the following results:

Hypothesis 1, which predicted greater lexical and constructional variation in the expression of GIVE in the subtitles than in the Bible translations, has been confirmed. The  $\kappa$  statistics of interrater agreement, which were used to measure the average agreement between the languages in encoding GIVE, show that the conceptualization is more cross-linguistically similar in the Bible translations than

in the subtitles, both at the constructional and the lexical levels. A quantitative comparison of proportions of canonical GIVE verbs and ditransitive constructions across the languages in the two subcorpora has shown that the GIVE exemplars from the Bible translations contain higher proportions of canonical GIVE verbs and ditransitive constructions than those exemplars that came from the subtitles. An MDS plot with the exemplars from two subcorpora has also revealed greater lexical variation of subtitle exemplars. Together, this evidence indicates that variation in translational practices and registers should be taken into account in future typological studies based on parallel corpora. To what extent this holds for different language subsystems (i.e. lexicon, morphology, syntax) is a question that requires further investigation.

Hypothesis 2 predicted greater agreement between the ten languages at the constructional level than at the lexical level. Again, the  $\kappa$  statistics based on the data in both subcorpora confirm this hypothesis. One possible explanation is that the ditransitive construction – understood very broadly as a combination of a verb of transfer and three related arguments – represents the event of GIVE iconically; in contrast, the lexemes are more arbitrary signs, and more prone to variation and change.

If Hypothesis 3 were true, one would expect to find a significant association between the lexical and constructional patterns. The significant  $\phi$  coefficients in all ten languages confirm this hypothesis, although the canonical GIVE verbs and the ditransitive constructions do not completely overlap.

Finally, Hypothesis 4 predicted that the genetically similar languages from the Germanic, Romance and Slavic groups, respectively, would be more similar to each other in expressing GIVE than to the languages from the other linguistic groups. This hypothesis was tested with the help of Multidimensional Scaling of the distances between the ten languages. The results only partially support Hypothesis 4, with some evidence of genetically interpretable clusters only at the lexical level. At the level of the ditransitive constructions, there are hardly any patterns that can be interpreted genetically. One possible explanation is that the more abstract patterns are more universal and less prone to historical changes. Three-dimensional plots featuring interpolation based on Kriging for the Germanic and Romance languages in the sample, as well as for two Slavic languages, reveal interesting differences in the prototypes of the canonical GIVE verbs for the three language groups.

This study was only one of the first attempts to do semantic typology based on parallel corpora, and many questions require further investigation. For example, this study employed binarized versions of the lexical and constructional variables

to permit simpler coding and easier interpretation. However, variation in the peripheral expressions of GIVE should be explored, as well. Next, as was mentioned in Section 3, it would be worthwhile to use lexemes from all languages, not just English, as inputs to search the exemplars of GIVE. Another intriguing direction for future investigation would be whether one can find areal patterns in the lexical and constructional categorization within the same language group (cf. van der Auwera et al. 2005), although such a study would require a large sample of related languages.

However, despite these limitations, the author would like to hope that she has managed to convince the reader that corpus-based semantic typology should become more multidimensional, both in terms of the text types and the abstraction level of the form-meaning pairings that cut semantic domains.

## References

- van der Auwera, Johan, Schalley, Ewa & Nuyts, Jan. 2005. Epistemic possibility in a Slavonic parallel corpus: A pilot study. In *Modality in Slavonic Languages. New Perspectives*, Björn Hansen & Petr Karlik (eds), 201–217. Munich: Sagner.
- Berlin, Brent & Kay, Paul. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley CA: University of California Press.
- Bresnan, Joan, Cueni, Anna, Nikitina, Tatiana & Baayen, R. Harald. 2007. Predicting the dative alternation. In *Cognitive Foundations of Interpretation*, Gerlof Bouma, Irene Kramer & Joost Zwarts (eds), 69–94. Amsterdam: Royal Netherlands Academy of Science.
- Colleman, Timothy. 2009. The semantic range of the Dutch double object construction: A colostruational perspective. *Constructions and Frames* 1: 190–221. DOI: 10.1075/cf.1.2.02col
- Cysouw, Michael & Forker, Diana. 2009. Reconstruction of morphosyntactic function: Non-spatial usage of spatial case marking in Tsezic. *Language* 85(3): 588–617. DOI: 10.1353/lan.0.0147
- Dixon, Robert M. W. 1973. The semantics of giving. In *The Formal Analysis of Natural Languages*, Maurice Gross, Morris Halle & Marcel-P. Schützenberger (eds), 205–223. The Hague: Mouton.
- Ellis, Nick & Ferreira-Junior, Fernando G. 2009. Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics* 7: 188–221. DOI: 10.1075/arcl.7.08ell
- Evans, Nicholas. 2010. Semantic typology. In *The Oxford Handbook of Linguistic Typology*, Jae Jung Song (ed.), 504–533. Oxford: OUP.
- Falissard, Bruno. 2012. psy: Various procedures used in psychometry. R package version 1.1. <<http://CRAN.R-project.org/package=psy>> (28 April 2014).
- Furrer, Reinhard, Nychka, Douglas & Sain, Stephan. 2013. fields: Tools for spatial data. R package version 6.8. <<http://CRAN.R-project.org/package=fields>> (28 April 2014)

- Geeraerts, Dirk. 1988. Where does prototypicality come from? In *Topics in Cognitive Linguistics*, [Current Issues in Linguistic Theory 50], Brygida Rudzka-Ostyn (ed.), 207–229. Amsterdam: John Benjamins. DOI: 10.1075/cilt.50.09gee
- Goldberg, Adele E. 2005. Argument realization: The role of constructions, lexical semantics and discourse factors. In *Construction Grammars: Cognitive Grounding and Theoretical Extensions*, Jan-Ola Östman & Miriam Fried (eds), 17–43. Amsterdam: John Benjamins. DOI: 10.1075/cal.3.03gol
- Goldberg, Adele E., Casenhiser, Devin & Sethuraman, Nitya. 2004. Learning argument structure generalizations. *Cognitive Linguistics* 14(3): 289–316.
- Haspelmath, Martin. 2011. Ditransitive constructions: The verb “Give.” In *The World Atlas of Language Structures Online*. Matthew S. Dryer & Martin Haspelmath (eds). Munich: Max Planck Digital Library. <<http://wals.info/chapter/105>> (7 November 2013).
- Janssen, Theo A. J. M. 1998. Giving in Dutch: An intra-lexematical and inter-lexematical description. In *The Linguistics of Giving* [Typological Studies in Language 36], John Newman (ed.), 267–306. Amsterdam: John Benjamins. DOI: 10.1075/tsl.36.10jan
- de Leeuw, Jan & Mair, Patrick. 2009. Multidimensional Scaling Using Majorization: SMACOF in R. *Journal of Statistical Software* 31(3): 1–30. <<http://www.jstatsoft.org/v31/i03/>> (28 April 2014).
- Maechler, Martin, Rousseeuw, Peter, Struyf, Anja, Hubert, Mia & Hornik, Kurt. 2013. cluster: Cluster Analysis Basics and Extensions. R package version 1.14.4.
- Malchukov, Andrej, Haspelmath, Martin & Comrie, Bernard. 2010. Ditransitive constructions: A typological overview. In *Studies in Ditransitive Constructions. A Comparative Handbook*, Andrej Malchukov, Martin Haspelmath & Bernard Comrie (eds), 1–64. Berlin: De Gruyter Mouton.
- Mayer, Thomas & Cysouw, Michael. In press. Creating a massively parallel Bible corpus. *Proceedings of LREC 2014, Reykjavik*, 26–31 May 2014.
- Meyer, David, Zeileis, Achim & Hornik, Kurt. 2012. vcd: Visualizing Categorical Data. R package version 1.2–13. <<http://CRAN.R-project.org/package=vcd>> (28 April 2014)
- Murphy, Gregory L. 2002. *The Big Book of Concepts*. Cambridge MA: The MIT Press.
- Nerlove, Sarah & Romney, Kimball. 1967. Sibling terminology and cross-sex behaviour. *American Anthropologist* 69: 179–187. DOI: 10.1525/aa.1967.69.2.02a00050
- Newman, John. 1996. *Give: A Cognitive Linguistic Study*. Berlin: Mouton de Gruyter.
- Newman, John (ed.). 1998a. *The Linguistics of Giving* [Typological Studies in Language 36]. Amsterdam: John Benjamins. DOI: 10.1075/tsl.36
- Newman, John. 1998b. The origin of the German es gibt construction. In Newman (ed.), 307–325.
- Och, Franz J. & Ney, Hermann. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics* 29(1): 19–51. DOI: 10.1162/089120103321337421
- R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <<http://www.R-project.org>> (28 April 2014).
- Roberts, John R. 1998. GIVE in Amele. In Newman (ed.), 1–33.
- Stefanowitsch, Anatol & Gries, Stefan T. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2): 209–243. DOI: 10.1075/ijcl.8.2.03ste
- de Vries, Lourens. 2009. Some remarks on the use of Bible translations as parallel texts in linguistic research. *STUF – Sprachtypologie und Universalienforschung* 60(2): 148–157.

- Wälchli, Bernhard. 2010. Similarity semantics and building probabilistic semantic maps from parallel texts. *Linguistic Discovery* 8(1): 331–371.
- Wälchli, Bernhard & Cysouw, Michael. 2012. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50(3): 671–710.  
DOI: 10.1515/ling-2012-0021
- von Waldenfels, Ruprecht. 2012. *The Grammaticalization of “give” + Infinitive: A Comparative Study of Russian, Polish and Czech*. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110293777

Uncorrected proofs - © John Benjamins Publishing Company

