Linguistic Issues in Language Technology – LiLT Submitted, January 2012

# The Meaning of Lexical Causatives in Cross-Linguistic Variation

Tanja Samardžić and Paola Merlo

Published by CSLI Publications

#### LiLT volume 7, issue 12

#### January 2012

# The Meaning of Lexical Causatives in Cross-Linguistic Variation

TANJA SAMARDŽIĆ AND PAOLA MERLO, University of Geneva

#### Abstract

The causative alternation has been recognised in the linguistic literature as one of the most widely spread linguistic phenomena, attested in almost all languages, although differently realised and involving partially different sets of verbs. In this paper, we identify the degree of spontaneity of the event described by a verb as a general component of meaning of alternating verbs which underlies the within-language and cross-linguistic variation in their realisations. We first establish that a corpus-based measure of this property, the ratio of the frequency of usage of the causative and anticausative form, is strongly correlated to an independent typological measure (Haspelmath, 1993). Then we examine the influence of this property on the cross-linguistic realisations of verbs. We find that the degree of variation and parallelism in forms across languages is strongly related to the degree of spontaneity of the verb.

LiLT Volume 7, Issue 12, January 2012. The Meaning of Lexical Causatives in Cross-Linguistic Variation. Copyright © 2012, CSLI Publications.

#### 1 Introduction

The term causative(/inchoative) alternation refers to the fact that some verbs can be realised in a sentence both as transitive (1a) and as intransitive (1b), where both versions express the same event, with the only difference being that the transitive version specifies the *causer* of the event (*Mary* in (1a)), and the intransitive version does not. The transitive version, thus, can be termed *causative* and the intransitive *anticausative*. The verbs that can participate in this alternation are commonly referred to as *lexical causatives*.

The causative alternation has been recognised in the linguistic literature as one of the most widely spread linguistic phenomena, attested in almost all languages (Schäfer, 2009). Most of the alternating verbs are lexical counterparts across many languages. Some verbs, however, can alternate in some languages, while their lexical counterparts in other languages cannot. Languages also differ in the way the alternation is realised: whether it is morphologically marked, what kind of marking is applied, which version, causative or anticausative, is marked in the languages where there is marking.

(1)a. Mary broke the vase.

b. The vase broke.

- (2)a. The children played.
  - b. \* The parents played the children.
- (3)a. The parents bought the toys.
  - b. \* The toys bought.
- (4)a. Their photo was hanging on the wall.
  - b. They were hanging their photo on the wall.
- (5)a. The baker cut the bread.
  - b. \* The bread cut.
- (6)a. The flowers suddenly bloomed.
  - b. \* The summer bloomed the flowers.

The most important generalisation about lexical causatives is that they describe an event involving a change of state of one of the participants (Levin and Rappaport Hovav, 1994). The alternations in (2) and (3) are not possible because *children* and *toys* are not understood as undergoing a change. This generalisation, however, does not explain all the cases of the alternation. Some verbs that alternate do not describe a change of state (4) and some verbs that do not alternate do ((5) and (6)).

To account for the cases illustrated in (4) and (6), Levin and Rappaport Hovav (1994) make a difference between "externally" and "internally" caused events. For a verb to be able to alternate, the causer of the event that it describes needs to be external to the changing entity. The alternation in (6) is not possible because the cause of *blooming* is inherent to *flowers*, which means internal to the event. On the other hand, external causation makes the alternations in (4) possible even though the participant (*photo*) is not undergoing a change.

For the case illustrated in (5) no generalisations are proposed. The fact that the alternation is not possible even though the participant *bread* is undergoing a change is explained by idiosyncratic lexical properties of the verb (*cut*) such that it requires expressing the causer (Levin and Rappaport Hovav, 1994). According to Haspelmath (1993), expressing the causer can become obligatory when the meaning of the verb is highly specified (e.g. *decapitate* in English).

Another observation for which no generalisations have been proposed is the difference in realisations of lexical causatives across languages, specifically, the fact that there are verbs that do alternate in some languages, while their counterparts in other languages do not. For example, Greek and Hindi counterparts of *kill* and *destroy* have intransitive versions (Alexiadou et al., 2006), while typically intransitive verbs of manner of motion (*run, walk, fly*) can have transitive versions in English, but not in French or German (Schäfer, 2009). A related question that has not been addressed in the linguistic literature at all is whether all the alternating verbs are equally likely to have different realisations across languages.

In this paper, we identify the degree of involvement of an external causer in an event described by a verb as a general component of meaning of lexical causatives which underlies all the observed behaviours of these verbs, including the cross-linguistic variation. We call this property the *degree of spontaneity*.

Following recently developed empirical approaches to the interface between syntax and semantics based on parallel treebanks (Dyvik et al., 2009), we study realisations of a wide range of alternating verbs extracted from syntactically annotated corpora. These corpora allow us to ask questions about the distribution of parallel instances, and to study grammatical systems comparatively, at the level of individual tokens.

We first establish a corpus-based measure of spontaneity for each verb: we show how the ratio of the frequency of usage of the causative and anticausative form in a corpus is an indicator of the degree of spontaneity established by an independent measure (Haspelmath, 1993). Having developed a corpus-based indicator of this property, we then examine its influence on the cross-linguistic realisations of verbs. We find that the degree of alternation and parallelism of forms across languages is strongly related to the degree of spontaneity of the verb.

#### 2 Experiment 1: The Scale of Spontaneous Occurrence

The degree of involvement of an external causer in the event described by a verb has already been recognised as a quantifiable property of meaning of lexical causatives. Studying the typology of morphological marking of 31 pairs of alternating verbs across a wide range of languages, Haspelmath (1993) notices that, despite the fact that languages apply different types of morphological marking, certain alternating verbs tend to get the same kind of marking across languages. Verbs such as lexical equivalents of English *freeze*, dry, melt tend to be marked when used causatively in many different languages, while the equivalents of English gather, open, break, close tend to be marked in their anticausative uses. Calculating the ratio of the number of languages that mark the anticausative version to those that mark the causative version for each pair of verbs, Haspelmath (1993) proposes a universal scale of *increasing likelihood of spontaneous occurrence*, shown in (7). The verbs with a low anticausative to causative ratio describe events that are likely to happen with no external causer involved. The use of such a verb with an expressed causer is morphologically marked in the majority of languages. The verbs with a high ratio typically specify an external causer. Their use with no causer specified tends to get some kind of morphological marking across languages.

 $\begin{array}{ll} (7) freeze > dry > melt > & \dots & > gather > open > break > close \\ & \text{low A/C (spontaneous)} & & \text{high A/C (non-spontaneous)} \end{array}$ 

In this interpretation, the cross-linguistic anticausative to causative morphological marking ratio is an observable and measurable indicator of a lexical property of verbs. It expresses the degree to which an external cause is involved in the event described by the verb.

The notion of spontaneous occurrence can be related to the distinction between internally and externally caused event argued for in the qualitative analyses. Both notions concern the same lexical property of verbs — the involvement of an external causer in the event described by a verb. The events that are placed on the spontaneous extreme of the scale would be those that can be perceived as internally caused. The occurrence of an external cause in these events is very unlikely. The externally caused events would correspond to a wider portion of the scale of spontaneous occurrence, including not just the events on the non-spontaneous extreme of the scale, but also those in the middle of the scale. The Meaning of Lexical Causatives in Cross-Linguistic Variation / 5

To determine the positioning of any given verb on the scale of spontaneous occurrence as proposed by Haspelmath (1993), it would be necessary to know the ratio of anticausative to causative morphological marking of its counterparts in many different languages. Since we want to take into account as many alternating verbs as possible, we need to represent this value in such a way that it can be calculated automatically for several hundreds of verbs.

To be able to do this, we explore the possibility of using a monolingual corpus-based measure as a representation of the positioning of the verb on the universal scale of spontaneous occurrence, suggested recently by Haspelmath (2008). We calculate the ratio between the corpus frequencies of causative (active transitive) and anticausative (intransitive) uses for those verbs for which the typological information is available: all the verbs listed by Haspelmath (1993) with exception of two verbs which were left out for technical reasons.<sup>1</sup> We then measure the strength of the correlation between the ranks obtained by the two measures.

We expect to find a correlation between the distribution of morphological marking on the verbs across languages and the distribution of their transitive to intransitive uses in a corpus, because it is well established that there is a correspondence between markedness and frequency. As marked forms are expected to be less frequent than unmarked forms, verbs that tend to have anticausative marking are expected to be used more as causative, and verbs that tend to have causative marking as anticausative. To obtain a positive, and not a negative, correlation, we calculate the C/A and not the A/C ratio as Haspelmath (1993).

#### 2.1 Materials and Methods

Transitive, intransitive, and passive instances of the verbs were extracted from the English side of the parallel corpus Europarl (Koehn, 2005), version 3, which contains around 1'500'000 sentences for each language.<sup>2</sup> The corpus was automatically parsed using the MaltParser, a data-driven system for building parsers for different languages (Nivre et al., 2007). Only the instances with all the arguments realised in the same clause were taken into account, excluding the instances that occur

 $<sup>^{1}</sup>$ The two verb pairs in question are those containing phrasal verbs, "go out/put out" and "get lost/lose". These verb pairs were left out of the calculation because extracting their instances from the corpus could not be done using methods already developed for extracting other verb instances.

 $<sup>^{2}</sup>$ We collect the data about active transitive and passive transitive uses separately because it is only the active transitive use that requires overt realisation of an external causer (as the subject).

with either the subject or the object moved or elided.<sup>3</sup>

To asses the strength of correlation between the corpus-based C/A ratio and the A/C ratio based on the typology of morphological marking on the verbs, we perform a correlation test between the rankings of the same verbs based on the two measures.

#### 2.2 Results

We obtain the correlation score r = 0.77, p < 0.01. With one outlier<sup>4</sup> removed, the score improves to r = 0.84, p < 0.01, suggesting a very good correlation between the two sources of data. Figure 1 shows the scattergram representing the correlation. The asymmetry in the scattergram seems to indicate that the corpus-based measure tends to rank events as less spontaneous than the typology-based measure.

Such a strong correlation between typological distribution of different morphological markings in lexical causatives and the distribution of transitive and intransitive uses of these verbs in a monolingual corpus suggests that the probability of occurrence of an external causer in an event described by a verb is a grammatically relevant component of the meaning of alternating verbs. It also suggests that the ratio of transitive to intransitive instances of a verb in a corpus can be used as a measure of the degree to which an external causer is involved in an event described by the verb.

#### 3 Experiment 2: Cross-linguistic Variation

In analysing cross-linguistic variation in realisations of lexical causatives, we try to determine whether a verb can be expected to have consistent or inconsistent realisations across languages depending on the degree to which an external causer is involved in the event described by the verb.

Studying instances of translations of lexical causatives in a parallel corpus makes it possible to control for any pragmatical and contextual factors that may be involved in a particular realisation of a lexical causative. Since translation is supposed to express the same meaning in the same context, we can assume that the same factors that influence a

<sup>&</sup>lt;sup>3</sup>The parser's error rate of 8% on these structures is assessed in a manual evaluation of 100 randomly chosen instances. Judging by the great diversity of the error types in our sample, we can expect the errors to be spread evenly over the extracted instances, having no major influence on the findings.

<sup>&</sup>lt;sup>4</sup>Verb *freeze* is frequently used in our corpus in its non-literal sense (e.g. *freeze* pensions, *freeze* assets), while the sense that was taken into account by Haspelmath (1993) is most likely the literal meaning of the verb (as in *The lake freezed*.). This is why the verb's corpus based ranking was very different from its typology based ranking.



FIGURE 1 The correlation between the rankings of verbs on the scale of spontaneous occurrence

particular realisation of a verb in a clause in one language influence the realisation of its translation in the corresponding clause in another language. Any potential differences in the form of the two parallel clauses should be explained by the lexical properties of the verbs or by structural differences between languages.

Unlike the previous research on lexical causatives that is either monolingual or typological, we take a micro-comparative approach and study the realisations of lexical causatives in English and German. These two languages are genetically and geographically very close which is why they can be expected to be very similar with respect to many aspects of their structure, including the realisations of lexical causatives. However, being two different languages, they are expected to be different in some respects.

By studying closely related languages, we reduce potential sources of cross-linguistic variation to the minimum, making the potential trends in the variation more apparent. In the case of our study, we can expect fewer verbs to be differently realised in English and German than it would be the case in two distant languages, with fewer potential sources of variation. On the other hand, if a verb is inconsistently realised in English and German, inconsistent realisations of this verb can be expected in any two languages. This approach is in line with some recent trends in theoretical linguistics centred around the idea that minimal abstract units of language structure can be isolated by identifying minimal structural divergence in two similar languages or even dialects of the same language (Kayne, 2005).

We perform a statistical analysis of a sample of parallel instances of lexical causatives in English and German, which we divide into three subsamples: expressions of spontaneous events, expressions of non-spontaneous events and expressions of the events that are neutral with respect to spontaneity. Given that spontaneity of an event correlates strongly with transitive or intransitive use monolingually, and given that translations are meaning-preserving, we expect to find an interaction between the level of spontaneity of the event described by the verb and its cross-linguistic syntactic realisation. Specifically, we expect syntactic realisations consistent with the lexical semantics of the verb to be carried across languages in a parallel fashion, while those that are inconsistent are expected to show a tendency towards the consistent realisation. For example, we expect intransitive realisations to stay intransitive, and transitives to be often transformed into intransitives when verbs describe spontaneous events. Since the probability of both realisations is similar in neutral instances, we expect to find fewer transformations than in the other two groups.

#### 3.1 Materials and Methods

The verbs included in our study are the 354 English verbs listed as alternating by Levin (1993). We calculate the ratio between the corpus frequencies of causative (active transitive) and anticausative (intransitive) uses in the same way as this was done for the verbs in Experiment 1 (Section 2.1). German translations of these instances were extracted from the German side of Europarl, syntactically parsed using the same system which was used for the English side of the corpus. The corpus was word-aligned using the system GIZA++ (Och and Ney, 2003). Both for the syntactic parses and word alignments, we reuse the data provided by Bouma et al. (2010) who used these tools to parse and align the English and the German side of Europarl.

German translations of instances of English lexical causatives are extracted in two steps. First, all verbs occurring as transitive, intransitive, and passive were extracted from German sentences that are sentencealigned with the English sentences containing the instances of lexical



FIGURE 2 Examples of parallel instances of lexical causatives.

FIGURE 3 Density distribution of the Sp value in the instances

density.default(x = lexbias)

causatives. As in the case of English, we extract only the instances of verbs with the arguments realised in the same clause. These instances were considered as candidate translations. The instances that are the translations of the English instances were then selected on the basis of word alignments. Instances where at least one element (the verb, the head of its object, or the head of its subject) is aligned with at least one element in the English instance were considered aligned.

This method allows extracting only translations with limited crosslinguistic variation, which corresponds well to our goal of minimising the variation. Only the instances of English verbs that are translated with a corresponding finite verbal form in German are extracted, excluding the cases where English verbs are translated to German with a corresponding non-finite form such as infinitive, nominalization, or participle in German.

The extracted parallel instances were then combined with the information on the corpus-based spontaneity score of the verb used in the instance, expressed as the logarithm of the ratio of transitive to intransitive uses of the verb (see Figure 2 for an illustration).

The three groups of instances are defined according to the density distribution of the Sp value. As it can be seen in Figure 3, roughly symmetric points of low density are around values -1 and 1. We consider the instances containing the verbs with the value of Sp inferior to -1

as the low value group. These are expressions of spontaneous events. Instances containing a verb with the Sp value superior to 1 are considered as belonging to the high value group, representing expressions of non-spontaneous events. The instances in between the two values are considered medium value instances. This division gives approximately symmetric groups with 3'107 instances with high Sp values, 2'822 instances with low Sp values, and 7'104 instances with medium Sp values.

#### 3.2 Results and Discussion

Table 1 shows the frequencies of the realisations of lexical causatives in parallel English and German instances for the whole sample of instances, as well as for the three sub-samples. The three most frequent combinations of forms in each group of parallel instances are highlighted to show the changes in the distribution of combinations of forms in the two languages across groups.

The overview of the frequencies suggests that lexical properties of verbs influence their cross-linguistic realisations.

The table that shows occurrences over the whole sample indicates that, both in English and in German, intransitives are more frequent than transitives, which are, in turn, more frequent than passives (marginal distributions). The non-parallel translations cover 32% of the cases.

When we partition the occurrences by the spontaneity of the event, the distribution changes, despite the fact that these are distributions in translations, and therefore subject to very strong pressure in favour of parallel constructions.

In the group of instances containing verbs that describe events around the middle of the scale of spontaneous occurrences, the parallel combinations are the most frequent, as in the distribution of the whole set, with an even more markedly uniform distribution (29% of nonparallel translations). This means that the verbs that describe events that are neither spontaneous or non-spontaneous tend to be used in the same form across languages. The probabilities of the two realisations are similar in these verbs, which means that they can be expected to occur with similar frequency across languages. Since both realisations are frequent in these verbs, they can be expected to alternate in the majority of languages.

The distribution of the forms is different in the groups of instances containing verbs that describe events on the extremes of the scale of spontaneous occurrence. The parallel realisations are frequent only for the forms that are consistent with the lexical properties (intransitive for spontaneous events and transitive for non-spontaneous events). An

WHOLE		German							
		Intransitive		Transitive		Passive		TOTAL	
SAMPLE		N	%	Ν	%	Ν	%	N	%
English	Intransitive	3504	<b>27</b>	1001	8	314	2	4819	37
	Transitive	1186	9	2792	<b>21</b>	369	3	4347	33
	Passive	781	6	517	4	2569	<b>20</b>	3867	30
	TOTAL	5471	42	4310	33	3252	25	13033100	
Spontaneous		German							
		Intransitive		Transitive		Passive		TOTAL	
	EVENTS	N	%	N	%	Ν	%	N	%
English	Intransitive	1733	61	495	17	102	4	2330	82
	Transitive	182	6	132	5	18	1	332	11
	Passive	35	1	23	1	102	4	160	6
	TOTAL	1950	68	650	23	222	9	2822	100
Non-		German							
SPONTANEOUS		Intransitive		Transitive		Passive		TOTAL	
5P	ONTANEOUS	intrans	sitive	Trans	itive	ras	sive	101	L'AL
5P	EVENTS	N N	%	Trans: N	%	r as N	sive %	N N	%
	EVENTS Intransitive	N 74	$\frac{\%}{2}$	Trans N 72	$\frac{11}{2}$	1 as N 29	$\frac{\frac{81}{8}}{1}$	101 N 175	$\frac{\%}{5}$
lish lsh	EVENTS Intransitive Transitive	N 74 288	$\frac{\frac{1}{2}}{9}$	1rans: N 72 948	2 31	125	$\frac{\frac{81}{6}}{\frac{1}{4}}$	N 175 1361	$\frac{\frac{\%}{5}}{44}$
Inglish	EVENTS Intransitive Transitive Passive	Intrans           N           74           288           448	14	1rans: N 72 948 289	10 10 10 10 10 10 10 10 10 10 10 10 10 1	125 834	8176 <u>%</u> <u>1</u> <u>4</u> <b>27</b>	N 175 1361 1571	$ \frac{\frac{1}{2}}{\frac{5}{51}} $
English	EVENTS Intransitive Transitive Passive TOTAL	Intrans           N           74           288           448           810	%           2           9           14           25	Trans:           N           72           948           289           1309	10 10 43	N 29 125 834 988	%           1           4           27           32	N 175 1361 1571 3107	$ \frac{\%}{5} $ $ \frac{44}{51} $ $ 100 $
English	EVENTS Intransitive Transitive Passive TOTAL	Intrans           N           74           288           448           810	% 2 9 14 25	N           72           948           289           1309	$ \frac{100}{43} $	N 29 125 834 988 an	%       1       4       27       32	N 175 1361 1571 3107	
English	EVENTS Intransitive Transitive Passive TOTAL NEUTRAL	Intrans N 288 448 810 Intrans	% 2 9 14 25 sitive	Trans           N           72           948           289           1309	$\frac{\%}{2}$ $\frac{31}{10}$ $\frac{43}{\text{Germ}}$ itive	N 29 125 834 988 an Pas	%           1           4           27           32	N 175 1361 1571 3107	%           5           44           51           100
English	EVENTS Intransitive Transitive Passive TOTAL NEUTRAL EVENTS	Intrans N 74 288 448 810 Intrans N	%       2       9       14       25       sitive       %	Trans           N           72           948           289           1309           Trans           N	$ \frac{\%}{2} $ $ \frac{31}{10} $ $ \frac{43}{6} $ $ \frac{31}{6} $ $ $	Pas           N           29           125           834           988           an           Pas           N	$\frac{\%}{1}$ $\frac{4}{27}$ $\frac{32}{32}$	N 175 1361 1571 3107 Tor N	%           5           44           51           100
h English	EVENTS Intransitive Transitive Passive TOTAL NEUTRAL EVENTS Intransitive	Intrans N 288 448 810 Intrans N 1697	%           2           9           14           25           sitive           %           24	Trans           N           72           948           289           1309           Trans           N           434	$ \frac{\frac{\%}{2}}{\frac{31}{10}} $ $ \frac{31}{6} $	Pas           N           29           125           834           988           an           Pas           N           183	$\frac{\%}{1}$ $\frac{1}{27}$ $\frac{32}{32}$ sive $\frac{\%}{3}$	N 175 1361 1571 3107 Tor N 2314	%           5           44           51           100           FAL           %           33
çlish English ç	EVENTS Intransitive Transitive Passive TOTAL NEUTRAL EVENTS Intransitive Transitive	Intrans N 288 448 810 Intrans N 1697 716	%           2           9           14           25           sittive           %           24           10	Trans           N           72           948           289           1309           Trans           N           434           1712	$ \frac{1}{31} $ $ \frac{1}{43} $ $ \frac{1}{31} $ $ \frac{1}{31} $ $ \frac{1}{31} $ $ \frac{1}{31} $ $ \frac{1}{32} $ $ $	Pas           N           29           125           834           988           an           Pas           N           183           226	$\frac{1}{32}$ sive $\frac{\%}{3}$ $3$	N 175 1361 1571 3107 N 2314 2654	$ \begin{array}{r} & & \\ & & \\ & & \\ \hline \hline & & \\ \hline \hline & & \\ \hline & & \\ \hline \hline & & \\ \hline \hline \\ \hline & & \\ \hline \hline \\ \hline & & \\ \hline \hline \hline \\ \hline \hline \hline \\ \hline \hline \hline \\ \hline \hline \hline \hline \\ \hline \hline$
Inglish English	EVENTS Intransitive Transitive Passive TOTAL NEUTRAL EVENTS Intransitive Transitive Passive	Intrans N 74 288 448 810 Intrans N 1697 716 298	%           2           9           14           25           sitive           %           24           10           4	Trans           N           72           948           289           1309           Trans           N           434           1712           205	$ \begin{array}{r} \text{itive} \\ & \\ \hline \\ \hline$	Pas           N           29           125           834           988           an           Pas           N           183           226           1633	sive % 1 4 27 32 sive % 3 3 23	N 175 1361 1571 3107 N 2314 2654 2136	$ \begin{array}{r} & & \\ & & \\ & & \\ \hline \hline & & \\ \hline \hline \\ \hline & & \\ \hline \hline & & \\ \hline \hline \\ \hline & & \\ \hline \hline \hline \\ \hline \hline \hline \\ \hline \hline \hline \hline \\ \hline \hline \hline \hline \hline \hline \hline \\ \hline \hline$

The Meaning of Lexical Causatives in Cross-Linguistic Variation / 11

 TABLE 1 Contingency tables for the English and German forms in different samples of parallel instances.

atypical instance of a verb in one language (e.g. transitive instance of a verb that describes a spontaneous event) is not preserved across languages. These realisations tend to be transformed into the typical form in another language. For example, German transitives are much less frequent in the spontaneous events group than in the non-spontaneous events group, while English intransitive non-spontaneous verbs are only 5% compared to 82% of the spontaneous group. The atypical realisations of these verbs are thus rare across languages, which means that they might be entirely absent in some languages. In the languages in which these realisations are found, the verbs alternate, while in the languages where these realisations are not found the verbs do not alternate. This means that the verbs describing events on the extremes of the scale of spontaneous occurrence can be expected to alternate in a smaller range of languages.

## 4 Related Work

Other empirical studies highlight important quantitative aspects of the causative alternation, showing that the frequencies of certain uses of verbs, an observable and measurable property, can be used as empirical evidence of their lexical properties. Merlo and Stevenson (2001) show that information collected from instances of verbs in a corpus can be used by an automatic classifier to distinguish between three verb classes: manner of motion verbs, which alternate only in a limited number of languages, externally caused change of state verbs, alternating across languages, and performance/creation verbs, which are not lexical causatives. McKoon and Macfarland (2000) provide corpus evidence for the distinction between externally and internally caused events. They then use this evidence to improve the classification of verbs proposed by Levin (1993).

Cross-linguistic empirical studies of realisations of causative verbs suggest that cross-linguistic realisations can also be influenced by specific grammatical properties of languages. Specifically, the work of Wolff et al. (2009) provides some evidences for the fact that transitive realisations of verbs are generally more likely in English than in other languages, including German. Our own study of the correlation between the universal scale of spontaneous occurrence and the use of English alternating verbs in a corpus seems to point to the same asymmetry. It can be noticed in Figure 1 that the rankings based on the corpus measure tend to be lower than the rankings based on the typological data. This means that the verbs in English are used causatively more than it could be expected on the basis of the position of the event that they describe on the universal scale of spontaneous occurrence.

# 5 Conclusions and Future Work

Our study proposes an approach to lexical causatives based on data collected from syntactically annotated parallel corpora. The analysis of the corpus realisations of these verbs provides evidence that the probability of occurrence of an external cause in the event described by a verb is a grammatically relevant lexical property. The cross-linguistic variation in realisations of lexical causatives is influenced by this property. Verbs that describe events on the extremes of the scale of spontaneous occurrence are more likely to have different realisations across languages than those that describe events in the middle of the scale.

Our data analysis, in accordance with other studies, indicates that grammatical differences between languages can be another factor underlying cross-linguistic variation in lexical causatives. The nature of the interaction between lexical properties of verbs and grammatical properties of languages still needs to be examined and explained both in a single language and cross-linguistically. This interaction can be related to the variation in morphological marking, such as, for example, the fact that the causative alternation is marked in some languages, but not in others, or the fact that, in languages in which the alternation is marked, some verb uses can be marked while others are not.

### Acknowledgments

The work described in this paper was financed by the SNSF Project No. 122643.

#### References

- Alexiadou, Artemis, Elena Anagnostopoulou, and Florian Schäfer. 2006. The properties of anticausatives crosslinguistically. In M. Frascarelli, ed., *Phases of Interpretation*, pages 187–212. Berlin, New York: Mouton de Gruyter.
- Bouma, Gerlof, Lilja Øvrelid, and Jonas Kuhn. 2010. Towards a large parallel corpus of cleft constructions. In Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10). Valletta, Malta: European Language Resources Association.
- Dyvik, Helge, Paul Meurer, Victoria Rosén, and Koenraad De Smedt. 2009. Linguistically motivated parallel parsebanks. In M. Passarotti, A. Przepiórkowski, S. Raynaud, and F. Van Eynde, eds., Proceedings of the Eight International Workshop on Treebanks and Linguistic Theories, pages 72-82. Milano: EDUCatt.
- Haspelmath, Martin. 1993. More on typology of inchoative/causative verb alternations. In B. Comrie and M. Polinsky, eds., *Causatives and tran*sitivity, vol. 23, pages 87–121. Amsterdam/Philadelphia: John Benjamins Publishing Co.
- Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. Cognitive Linguistics 19(1):1-33.
- Kayne, Richard. 2005. The Oxford Handbook of Comparative Syntax, chap. Some notes on comparative syntax, with special reference to English and French. Oxford University Press.
- Koehn, Philipp. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of MT Summit 2005*. Phuket, Thailand.

- Levin, Beth. 1993. English verb classes and alternations : a preliminary investigation. Chicago: The University of Chicago Press.
- Levin, Beth and Malka Rappaport Hovav. 1994. A preliminary analysis of causative verbs in English. *Lingua* 92:35-77.
- McKoon, Gail and Talke Macfarland. 2000. Externally and internally caused change of state verbs. *Language* 76(4):833–858.
- Merlo, Paola and Susanne Stevenson. 2001. Automatic verb classification based on statistical distribution of argument structure. *Computational Linguistics* 27(3):373-408.
- Nivre, Joakim, Johan Hall, Jens Nilsson, Chanev Atanas, Gülşen Eryiğit, Sandra Kübler, Svetoslav Marinov, and Erwin Marsi. 2007. MaltParser: A language-independent system for data-driven dependency parsing. Natural Language Engineering 13(2):95-135.
- Och, Franz Josef and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics* 29(1):19–52.
- Schäfer, Florian. 2009. The causative alternation. In Language and Linguistics Compass, vol. 3, pages 641–681. Blackwell Publishing.
- Wolff, Phillip, Ga-Hyun Jeon, and Yu Li. 2009. Causal agents in English, Korean and Chinese: The role of internal and external causation. Language and Cognition 1(2):165-194.