4FX: Light Verb Constructions in a Multilingual Parallel Corpus

Anita Rácz[1], **István Nagy T.**[1], **Veronika Vincze**[2]
[1]Department of Informatics, University of Szeged
raczanita89@gmail.com, nistvan@inf.u-szeged.hu
[2]Hungarian Academy of Sciences, Research Group on Artificial Intelligence
vinczev@inf.u-szeged.hu

Abstract

In this paper, we describe 4FX, a quadrilingual (English–Spanish–German–Hungarian) parallel corpus annotated for light verb constructions. We present the annotation process, and report statistical data on the frequency of LVCs in each language. We also offer inter-annotator agreement rates and we highlight some interesting facts and tendencies on the basis of comparing multilingual data from the four corpora. According to the frequency of LVC categories and the calculated Kendalls coefficient for the four corpora, we found that Spanish and German are very similar to each other, Hungarian is also similar to both, but German differs from all these three. The qualitative and quantitative data analysis might prove useful in theoretical linguistic research for all the four languages. Moreover, the corpus will be an excellent testbed for the development and evaluation of machine learning based methods aiming at extracting or identifying light verb constructions in these four languages.

Keywords: light verb constructions, parallel corpus, multilinguality, English, Spanish, German, Hungarian

1. Introduction

Multiword expressions (MWEs) are lexical items that contain space or "idiosyncratic interpretations that cross word boundaries". They can be decomposed into single words and display lexical, syntactic, semantic, pragmatic and/or statistical idiosyncrasy (Sag et al., 2002; Calzolari et al., 2002; Kim, 2008). One subclass of MWEs are light verb constructions (LVCs). They are formed by the combination of a nominal and a verbal component where the noun is usually taken in one of its literal senses but the verb loses its original sense to some extent. Due of their idiosyncratic behavior, they often pose a problem to natural language processing (NLP) systems. For instance, in machine translation they cannot be directly translated as the verbal component of the same light verb constructions may differ from language to language. Here we offer some English, German, Spanish and Hungarian LVCs:

> *to have a walk – eine Spaziergang machen* (lit. a walk make) – *dar un paseo* (lit. give a walk) – *sétát tesz* (lit. walk-ACC make)
> *to reach an agreement – llegar a un acuerdo* (lit. arrive to an agreement) – *eine Absprache treffen* (lit. an agreement meet) – *megegyezésre jut* (lit. agreement-SUB get)

Here we describe 4FX, a quadrilingual (English–Spanish–German–Hungarian) parallel corpus annotated for light verb constructions. We present the annotation process and report statistical data on the frequency of LVCs in each language. We hope that the corpus will enhance multilingual research on light verb constructions both from a theoretical linguistic point of view and from a computational linguistic point of view (especially for the development of applications).

The structure of the paper is as follows. First, related corpora and related work on the NLP treatment of multiword expressions are presented. Then the corpus is described together with annotation principles and inter-annotator agreement rates are also provided. After presenting some statistical data on the corpus the paper concludes with illustrating how the corpus and the database can be exploited in several fields of NLP.

2. Related work

Annotated corpora of light verb constructions are essential in the automatic detection of light verb constructions. On the other hand, they may be exploited in theoretical linguistic research as well. We are aware of the following monolingual resources manually annotated for light verb constructions. Kaalep and Muischnek (2006; 2008) presented an Estonian database and a corpus of multiword verbs. Krenn (2008) reported a database of German PP-verb combinations. The Prague Dependency Treebank was also annotated for multiword expressions (Bejcek and Stranák, 2010), thus for light verb constructions too (Cinková and Kolářová, 2005). NomBank (Meyers et al., 2004) contains the argument structure of common nouns, including those occurring in support verb constructions as well. The VNC-Tokens dataset (Cook et al., 2008) contains annotated examples of literal and idiomatic uses of English verb + noun combinations. In the Wiki50 corpus several types of English multiword expressions (including LVCs) are annotated (Vincze et al., 2011). The corpus used in the experiments of Tu and Roth (2011) contains English light verb constructions. Tan et al. (2006) reports their results on corpus-based identification of light verb constructions in English. As for Hungarian, an annotated corpus and a database containing LVCs are described in Vincze and Csirik (2010). Previously, we created the SzegedParalellFX English–Hungarian parallel corpus, which is manually annotated for LVCs (Vincze, 2012). To the best of our knowledge, this is the only parallel corpus annotated for LVCs. In this work, we would like to extend this research track, which manifests in the creation of a quadrilingual paralell corpus annotated for LVCs.

3. The corpus

The JRC-Acquis Multilingual Parallel Corpus consists of legislative texts for a range of languages used in the European Union (Steinberger et al., 2006). For an earlier study on LVC detection (Vincze et al., 2013), we randomly selected 60 documents from the English version of the corpus and annotated LVCs in them. In this work, we annotate the Spanish, German and Hungarian equivalents of those 60 documents, thus yielding a quadrilingual parallel corpus named 4FX. It is important to emphasize, however, that the corpora are aligned only at the sentence level and not at the level of LVCs.

Data on annotated texts can be seen in Table 1.

	en	de	es	hu	Total
Sentences	5143	5,527	5,675	4,568	20,913
Tokens	94,747	89,523	107,851	92,707	384,828
Token/sent.	18.42	16.19	19.01	20.29	18.41

Table 1: Statistical data on the 4FX corpus.

As the table demonstrates, the English corpus of more than 94000 tokens and its parallel equivalents in the three other languages formed the basis of the manual annotation. Regarding the number of tokens, the Hungarian and English corpora are close to each other, but the number of Spanish tokens exceeds them by around 13 percent, while German falls behind by approximately 6 percent. Comparing the number of tokens and sentences, less obvious tendencies can be observed. Concerning the average sentence length, German occupies the last place, being Spanish and Hungarian in the middle and Hungarian on top.

3.1. Types of light verb constructions

As already described in Vincze (2012), light verb constructions may occur in various surface forms due to their syntactic flexibility. For the sake of simplicity, we give English examples here but these can be generalized for the other languages as well.

Besides the prototypical verb + noun combination (VERB), light verb constructions may be present in different syntactic structures, that is, in participles (PART, e.g. *photos taken*) and they may also undergo nominalization, yielding a nominal compound (NOM, e.g. *service provider*). We also distinctively marked split light verb constructions (SPLIT, e.g. ***a decision*** *has been recently* ***made***), where the noun and the verb are not adjacent in the sentence, which is especially frequent in German due to word order constraints. All the above types are annotated in the corpus texts since they occur relatively frequently in each language (see Table 3).

3.2. Annotation principles

Two native speakers of Hungarian who could speak English, German and Spanish at an advanced level carried out the annotation. Corpus texts contain single annotation, i.e. one annotator worked on each text.

In order to annotate LVCs in different languages as uniformly as possible, we adapted the guidelines used during the construction of SzegedParalellFX. Thus, the test battery including questions such as *Can a verb (derived from the same root as the nominal component) substitute the construction?*, *When omitting the verb (e.g. in a possessive construction), can the original action be reconstructed?*, *Can the construction itself be nominalized?*, *Can the construction be passivized?* etc. was adapted for German and Spanish too. It should be noted that while in German linguistic traditions, constructions where the nominal component is the subject are not traditionally considered to be *Funktionsverbgefüge*, which is the German equivalent of the term *light verb construction*, here we marked them as LVCs in accordance with the other languages, for example: *Am 18. Juli 2005 fand eine mündliche Anhörung statt.* "An oral hearing was held on 18 July 2005."

Another language specific annotation principle was that we also annotated German LVCs where the nominal component was in the genitive case in case the meaning of the construction was to express an opinion, e.g. *der Ansicht/der Meinung sein* "to be of the opinion".

Complex predicates required special treatment in all the four languages. In such cases, we decided to mark only the main verb hence auxiliaries were not marked. The following German and English examples below illustrate this, which are translational equivalents:

Eine ***Entscheidung*** *ist* ***getroffen*** *worden.*

A ***decision*** *was* ***made****.*

Nominalized constructions were annotated regardless of whether they consist of one or even more elements, for example *szerződéskötés* "making a contract" in Hungarian or *Durchführung einer Untersuchung* "carrying out an investigation" in German.

With respect to prepositional LVCs, the preposition was marked as part of the nominal component. Moreover, German light verbs with separable prefixes required special and uniform treatment too because due to word order reasons, the prefix may occur in the last position of the sentence, separated from the verb it belongs to. In such cases, we decided to mark the separated prefixes again like verbs and at the annotation level, we had two verbal elements marked as part of the LVC.

3.3. Inter-annotator agreement rates

In order to measure the inter-annotator agreement rate, we randomly selected 10 documents in the four languages to be annotated by a second annotator as well. For dissimilar annotations, the two annotators discussed each case and their final decision was included in the gold standard data. Table 2 shows the inter-annotator agreement rates as compared to the gold standard annotation and for most of the cases, the level of agreement can be considered as substantively good.

Contrasting the κ-measures it is salient that the two annotators reached quite similar results on the Hungarian corpus. This is most probably due to the fact that they were annotating in their mother tongue. Annotator 1 achieved outstanding results on German and Spanish texts, while Annotator 2 reached higher rates on the English corpus. This might be explained by the fact that they had deeper knowledge of these languages and worked more often with them than with the rest of languages.

ENGLISH		**Precision**	**Recall**	**F-score**	**κ-measure**
GS vs. Annotator 1	VERB	81.39	83.33	82.35	71.07
	PART	84.09	82.22	83.15	71.52
	NOM	–	–	–	–
	SPLIT	36.63	0.5	42.11	36.72
	Unified	85.71	88.42	87.05	65.29
GS vs. Annotator 2	VERB	69.76	100.0	82.19	72.15
	PART	61.36	100.0	76.05	63.64
	NOM	–	–	–	–
	SPLIT	45.46	100.0	62.5	59.67
	Unified	63.26	100.0	77.5	75.52
GERMAN		**Precision**	**Recall**	**F-score**	**κ-measure**
GS vs. Annotator 1	VERB	75.0	92.31	82.75	79.87
	PART	100.0	94.73	97.29	96.68
	NOM	90.91	100.0	95.23	93.98
	SPLIT	80.0	91.42	85.33	76.59
	Unified	86.45	95.40	90.71	78.32
GS vs. Annotator 2	VERB	81.25	61.91	70.27	64.97
	PART	72.22	81.25	76.47	72.61
	NOM	86.36	95.0	90.47	88.46
	SPLIT	90.0	75.0	81.81	71.43
	Unified	84.38	77.14	80.59	63.81
SPANISH		**Precision**	**Recall**	**F-score**	**κ-measure**
GS vs. Annotator 1	VERB	94.23	89.09	91.58	78.93
	PART	90.0	85.71	87.81	84.16
	NOM	–	–	–	–
	SPLIT	85.71	85.71	85.71	84.49
	Unified	92. 40	87.95	90.12	81.93
GS vs. Annotator 2	VERB	59.61	88.57	71.26	42.05
	PART	25.0	83.33	38.46	30.76
	NOM	–	–	–	–
	SPLIT	28.57	1.0	44.45	42.28
	Unified	49.37	90.69	63.93	47.94
HUNGARIAN		**Precision**	**Recall**	**F-score**	**κ-measure**
GS vs. Annotator 1	VERB	86.45	96.22	91.07	85.08
	PART	78.85	93.18	85.42	77.81
	NOM	80.0	100.0	88.88	87.71
	SPLIT	28.57	40.0	33.33	30.42
	Unified	81.21	94.74	87.44	73.99
GS vs. Annotator 2	VERB	79.66	100.0	88.67	81.34
	PART	84.62	89.79	87.13	79.26
	NOM	66.66	100.0	80.0	78.01
	SPLIT	100.0	100.0	100.0	100.0
	Unified	84.96	100.0	91.87	75.54

Table 2: Inter-annotator agreement rates on the 4FX corpus

3.4. Statistics on corpus data

The total number and the number of the subtypes of light verb constructions in each language are presented in Table 3.

In Table 4, the number of LVCs is contrasted to the number of LVC lemmas and the frequency of each lemma on average is also presented. The number of hapax legomena (i.e. LVCs or light verbs that occur only once in the corpus) and their rate is also given here.

Tables 5 and 6 list the most frequent LVCs and light verbs in each language.

4. Comparing multilingual data

The comparison of the data on the four languages reveals interesting facts. First of all, it is salient that the number of light verb constructions in the languages are not the same: Hungarian texts seem to abound in LVCs while in English, there are about two third of the Hungarian frequency, German and Spanish being in the middle. However, further annotated corpora are needed, preferably from other domains, in order to see whether this difference in frequency is a specificity of the legal domain or it is a general characteristics of the languages.

Another interesting observation is that in German, there are

English		German		Spanish		Hungarian	
LVC	**#**	**LVC**	**#**	**LVC**	**#**	**LVC**	**#**
have regard	91	Hilfe gewähren “grant aid”	51	tener en cuenta “take into account”	79	támogatást nyújt “grant support”	89
enter into force	42	in Kraft treten “enter into force”	49	conceder ayuda “grant aid”	67	figyelembe vesz “take into account”	74
grant aid	38	Stellung nehmen “adopt a position”	46	entrar en vigor “enter into force”	45	részt vesz “take part”	59
take into account	32	Flug durchführen “operate a flight”	35	adoptar una medida “adopt measures”	27	hatályba lép “enter into force”	45
receive aid	20	Antrag stellen “hand in an application”	23	beneficiarse de una ayuda “receive aid”	20	döntést hoz “make a decision”	27
take account	18	Bezug nehmen “make reference”	20	efectuar un vuelo “operate a flight”	20	megállapodást köt “make a contract”	25
lay down a rule	12	Rechnung tragen “take account”	20	celebrar un acuerdo “conclude an agreement”	19	rendelkezésre áll “be at his disposal”	24
take measures	12	Lizenz erteilen “grant a licence”	14	poner en el mercado “place on the market”	14	hatást gyakorol “have impact”	18
impose an obligation	11	in Verkehr bringen “place on the market”	14	prestar un servicio “provide a service”	14	kérelmet benyújt “hand in an application”	16
meet a requirement	11	Ermäßigung gewähren “grant a reduction”	12	recibir una autorización “receive authorization”	12	támogatásban részesül “receive support”	16

Table 5: The most frequent LVCs in the 4FX corpus.

English		German		Spanish		Hungarian	
Light verb	**#**	**Light verb**	**#**	**Light verb**	**#**	**Light verb**	**#**
have	105	gewähren “guarantee’	87	tener “have”	150	vesz “take”	152
take	105	durchführen “execute”	78	conceder “grant”	94	nyújt “offer”	120
make	73	nehmen “take”	69	adoptar “adopt”	53	hoz “bring”	65
enter	46	treten “enter”	52	efectuar “effect”	50	tesz “make, put”	56
carry out	43	stellen “put”	32	entrar “enter”	46	kerül “get done”	54
grant	42	tragen “hold”	32	llevar “hold”	45	lép “step”	53
give	35	haben “have”	31	poner “put”	43	folytat “execute”	44
lay down	35	vornehmen “carry out”	25	realizar “realize”	41	benyújt “hand in”	43
meet	29	bringen “bring”	24	presentar “present”	35	végez “carry out”	43
receive	27	stehen “stand”	20	hacer “do, make”	33	ad “give”	41

Table 6: The most frequent light verbs in the 4FX corpus.

a lot more split constructions than in other languages. This is most probably due to the German word order: in subordinate clauses, it is the verb that is the last element of the clause thus it may happen that the nominal component of the light verb construction precedes the verb and they are not adjacent such as in (we provide the English equivalent as well):

[...] können sie gemäß Artikel 19 der Richtlinie ***einen Antrag*** *an die Kommission* ***richten***.

“they may **submit a request** to the Commission in accordance with Article 19 of the Directive.”

We also calculated Kendall’s coefficient for the four corpora, which reflects similarities among languages, concerning the frequency of LVC categories. According to the data, Spanish and German are very similar to each other (the coefficient is 1.0), Hungarian is also similar to both (0.9), but German differs from all these three to a greater degree (Kendall’s coefficient being 0.5 for Spanish and English and 0.3 for Hungarian). This may be another consequence of the German word order rules, which may be responsible for the bigger number of split constructions.

The number of LVC lemmas is the highest in Spanish and the number of light verbs is the highest in German. In English, both numbers are the lowest, which suggests that LVCs are less diverse in English than in the other languages, at least in the legal domain. The number of hapax

	en	de	es	hu	Total
NOM	37	151	82	199	469
	5.50%	18.73%	8.74%	6.91%	13.49%
VERB	245	265	519	384	1413
	36.40%	32.88%	55.33%	51.51%	40.65%
SPLIT	127	278	132	94	631
	18.87 %	34.49%	14.07%	8.88%	18.15%
PART	264	112	205	382	963
	39.23	13.90%	21.86%	36.07%	27.70%
All	673	806	938	1059	3476
	100.00	100.00	100.00	100.00	100.00

Table 3: Subtypes of light verb constructions in the 4FX corpus. **NOM:** nominal light verb constructions. **VERB:** verbal occurrences. **SPLIT:** split light verb constructions. **PART:** participial light verb constructions.

	en	de	es	hu
LVCs	678	806	938	1062
LVC lemmas	195	272	349	299
Average occurrence	3.48	2.96	2.69	3.55
LVC verbs	42	96	78	80
Average occ. in lemmas	4.64	2.83	4.47	3.74
Hapax LVCs	108	162	222	176
%	55.38	59.56	63.61	58.86
Hapax LVC verbs	11	35	28	24
%	26.19	36.46	35.90	30.00

Table 4: Statistics on the frequency of LVCs and LVC lemmas in the 4FX corpus.

LVCs and light verbs also reflects a similar picture, which might be of interest in the automatic detection of LVCs: a dictionary lookup method can probably achieve better results in English than is the other languages (Rácz et al., 2014).

As for a qualitative analysis of the data, it can be observed that there are some common LVCs that are frequent in each of the four languages such as:

to enter into force – in Kraft treten – entrar en vigor – hatályba lép

to grant aid – Hilfe gewähren – conceder ayuda – támogatást nyújt

Other construction occur among the top 10 LVCs in three of the languages (except for German) such as:

to take into account – tener en cuenta – figyelembe vesz

to receive aid – beneficiarse de una ayuda – támogatásban részesül

These are among the most frequent light verb constructions and they are also typical of the legal language. On the other hand, there are also language-specific light verb constructions in the data, which do not have an equivalent in all or any of the other languages just like the English phrase *having regard to* corresponds to the Hungarian phrase *tekintettel* regard-INS "with regard to".

If the most frequent light verbs are analyzed, we can again find some verbs that occur among the top 10 verbs in at least three languages, which are listed below:

- *take*, *nehmen* and *vesz*;
- *enter*, *treten* and *entrar*;
- *make*, *hacer* and *tesz*;
- *have*, *haben* and *tener*.

It is interesting to observe that while the verbs meaning "to make" are very frequent in a light verb construction in English, Spanish and Hungarian, the verb *machen* rarely occurs in German LVCs. On the other hand, there is no translational equivalent of the verb "to have" in Hungarian – possessive sentences like *I have a car* are expressed with the combination of a copula and some possessive suffixes on the noun –, which explains why no such verb occurs in the Hungarian data.

5. Conclusions

In this paper, we presented 4FX, a quadrilingual parallel corpus annotated for light verb constructions. We explained the theoretical basis of the annotation by describing the types of LVCs and the most essential annotation principles we followed. We provided statistical data on the corpus, we offered inter-annotator agreement rates too, and we highlighted some interesting facts and tendencies on the basis of comparing multilingual data from the four corpora.

The corpus contains 673 LVCs in English, 806 in German, 938 in Spanish and 1059 in Hungarian. The qualitative and quantitative data analysis might prove useful in theoretical linguistic research for all the four languages. Moreover, the corpus will be an excellent testbed for the development and evaluation of machine learning algorithms aiming at detecting light verb constructions in these four languages, which we would like to implement in the future.

The annotated corpus is available free of charge for research and educational purposes at our website: `http://www.inf.u-szeged.hu/rgai/mwe`.

6. Acknowledgments

István Nagy T. was funded by the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP-4.2.4.A/ 2-11/1-2012-0001 "National Excellence Program". The other authors were funded in part by the European Union and the European Social Fund through the project FuturICT.hu (grant no.: TÁMOP-4.2.2.C-11/1/KONV-2012-0013).

7. References

Bejcek, E. and Stranák, P. (2010). Annotation of multiword expressions in the Prague Dependency Treebank. *Language Resources and Evaluation*, 44(1-2):7–21.

Calzolari, N., Fillmore, C., Grishman, R., Ide, N., Lenci, A., MacLeod, C., and Zampolli, A. (2002). Towards best practice for multiword expressions in computational lexicons. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, pages 1934–1940, Las Palmas.

Cinková, S. and Kolářová, V. (2005). Nouns as Components of Support Verb Constructions in the Prague Dependency Treebank. In Šimková, M., editor, *Insight into Slovak and Czech Corpus Linguistics*, pages 113–139. Veda Bratislava, Slovakia.

Cook, P., Fazly, A., and Stevenson, S. (2008). The VNC-Tokens Dataset. In *Proceedings of the LREC Workshop Towards a Shared Task for Multiword Expressions (MWE 2008)*, pages 19–22, Marrakech, Morocco.

Kaalep, H.-J. and Muischnek, K. (2006). Multi-Word Verbs in a Flective Language: The Case of Estonian. In *Proceedings of the EACL Workshop on Multi-Word Expressions in a Multilingual Contexts*, pages 57–64, Trento, Italy. ACL.

Kaalep, H.-J. and Muischnek, K. (2008). Multi-Word Verbs of Estonian: a Database and a Corpus. In *Proceedings of the LREC Workshop Towards a Shared Task for Multiword Expressions (MWE 2008)*, pages 23–26, Marrakech, Morocco.

Kim, S. N. (2008). *Statistical Modeling of Multiword Expressions*. Ph.D. thesis, University of Melbourne, Melbourne.

Krenn, B. (2008). Description of Evaluation Resource – German PP-verb data. In *Proceedings of the LREC Workshop Towards a Shared Task for Multiword Expressions (MWE 2008)*, pages 7–10, Marrakech, Morocco.

Meyers, A., Reeves, R., Macleod, C., Szekely, R., Zielinska, V., Young, B., and Grishman, R. (2004). The NomBank Project: An Interim Report. In Meyers, A., editor, *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation*, pages 24–31, Boston, Massachusetts, USA, May 2 - May 7. ACL.

Rácz, A., Nagy T., I., and Vincze, V. (2014). 4FX: félig kompozicionális szerkezetek automatikus azonosítása többnyelvű korpuszon. In *MSzNy 2014 – X. Magyar Számítógépes Nyelvészeti Konferencia*, pages 317–324, Szeged, Hungary. University of Szeged.

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D. (2002). Multiword Expressions: A Pain in the Neck for NLP. In *Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002*, pages 1–15, Mexico City, Mexico.

Steinberger, R., Pouliquen, B., Widiger, A., Ignat, C., Erjavec, T., and Tufiş, D. (2006). The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006*, pages 2142–2147.

Tan, Y. F., Kan, M.-Y., and Cui, H. (2006). Extending corpus-based identification of light verb constructions using a supervised learning framework. In *Proceedings of the EACL Workshop on Multi-Word Expressions in a Multilingual Contexts*, pages 49–56, Trento, Italy. ACL.

Tu, Y. and Roth, D. (2011). Learning English Light Verb Constructions: Contextual or Statistical. In *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World*, pages 31–39, Portland, Oregon, USA. ACL.

Vincze, V. and Csirik, J. (2010). Hungarian corpus of light verb constructions. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1110–1118, Beijing, China. Coling 2010 Organizing Committee.

Vincze, V., Nagy T., I., and Berend, G. (2011). Multiword Expressions and Named Entities in the Wiki50 Corpus. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, pages 289–295, Hissar, Bulgaria, September. RANLP 2011 Organising Committee.

Vincze, V., Nagy T., I., and Zsibrita, J. (2013). Learning to detect English and Hungarian light verb constructions. *ACM Transactions on Speech and Language Processing (TSLP)*, 10(2), June.

Vincze, V. (2012). Light Verb Constructions in the SzegedParalellFX English–Hungarian Parallel Corpus. In *Proceedings of LREC 2012*, Istanbul, Turkey.