

On the machine translatability of semi-compositional constructions

Veronika Vincze

University of Szeged, Doctoral School in Linguistics, 6722 Szeged, Egyetem u. 2., Hungary
and University of Szeged, Department of Informatics, Human Language Technology Group,
6722 Szeged, Árpád tér 2., Hungary
vinczev@inf.u-szeged.hu

Abstract In this paper, some possible ways of translating semi-compositional constructions with the help of a computer are presented. Two possible methods are proposed: first, the one based on lexical functions, and, second, the statistical-based one making use of lexico-semantic relations between the noun and the verb. The application of both methods in machine translation would yield higher accuracy of translation. However, both methods require previous theoretical research, the results of which could be fruitfully applied in the field of computational linguistics, especially in machine translation.

1 Introduction

In this paper, I explore the possible ways of translating Hungarian noun + verb constructions with the help of computers. These constructions do not form a unified category, since, on the one hand, productive structures (such as *újságot olvas* newspaper-ACC reads 'to read a newspaper' or *levelet ír* letter-ACC writes 'to write a letter'), and, on the other hand, idiomatic expressions (such as *csütörtököt mond* lit. Thursday-ACC says 'it fails to work', *lépre csal* lit. comb-SUB entices 'to toll') can also be found. However, besides these constructions, there exist some expressions that are neither productive nor idiomatic and whose meaning is not totally compositional. For this latter type, examples from different languages are shown in (1)-(7). Since their meaning is the same, only literal translations are provided:

- (1) Hungarian: *előadást tart* presentation-ACC holds, *virágba borul* bloom-ILL falls, *lehetőség nyílik* possibility opens
- (2) English: *to give a lecture*, *to come into bloom*, *a possibility emerges*
- (3) German: *halten eine Vorlesung* to hold a presentation, *in Blüte stehen* in bloom to stand, *es gibt eine Möglichkeit* it gives a possibility
- (4) French: *faire une présentation* to make a presentation, *être en fleur* to be in bloom, *l'occasion se présente* the possibility itself presents
- (5) Portuguese: *fazer uma conferência* to make a presentation, *lançar flores* to throw flowers, *se oferecer ocasião* itself to present possibility
- (6) Italian: *tenere una conferenza* to hold a presentation, *essere in fiore* to be in bloom, *emerge la possibilità* emerges the possibility

- (7) Russian: *čitat' doklad* to read presentation, *pokryt'sja cvetami* to cover itself with bloom, *predstavljačetsja vozmožnost'* to emerge itself possibility

Several terms have been used for these constructions in the literature (Dobos 1991, 2001, Langer 2005). The most common ones are as follows: in German, they are called *Funktionsverbgefüge* (function verb constructions), in English, *complex verb structures*, *support verb constructions* or *light verb constructions* can be found, in French, *constructions à verbe support* (support verb constructions) is usually used, while we can find terms like *opisatel'nye vyraženiya* (descriptive expressions) in Russian, *costruzioni a verbo supporto* (support verb constructions) in Italian, *construções com verbo suporte* (support verb constructions) or *construções com verbo leve* (light verb constructions) in Portuguese.

In Hungarian, these constructions also have got several names: *körülíró szerkezetek* (periphrastic constructions) in (Sziklai 1986), *leíró kifejezések* (descriptive expressions) in (Dobos 1991), and *funkcióigés szerkezetek* (function verb constructions) following (Keszler 1992), however, the somewhat pejorative term *terpeszkedő szerkezetek* (“sprawling” constructions) occurs in the Hungarian Purists' Dictionary (Grétsy and Kemény 1996: 571) and in recent specialized articles as well (for instance, Heltai and Gósy (2005) focuses on the effects of sprawling constructions to linguistic processing).

As it can be seen from the examples given above, most names used for these constructions contain only one component of the construction, namely, the verbal component, suggesting that it is the verbal component that forms the head of the construction. However, since the verbal component functions only as the syntactic head of the construction – the nominal component being the semantic head of the construction (Dobos 2001) –, it is perhaps better not to include any of the two components in the name of the construction. That is why I will use the term semi-compositional constructions in my paper, following Langer (2005).

The aim of this paper is to overview the machine translatability of semi-compositional constructions. First, the characteristics of semi-compositional constructions will be examined in detail, then problems concerning the machine translatability of collocations in general will be presented, and some of the solutions offered for these problems will be shown. Finally, the way lexical functions can help the machine translatability of semi-compositional constructions will be demonstrated.

2 On the status of Hungarian bare common noun + verb constructions

In this section, I will examine the characteristics of Hungarian bare common noun + verb constructions, and I will present a possible classification of these constructions.

2.1 Earlier research on bare common noun + verb constructions

Bare common noun + verb constructions have been given some attention for a long time in Hungarian linguistic research. For instance, Komlósy (1992) classifies these constructions into four groups. The first group contains idioms or idiom-like expressions, whose meaning cannot be calculated from meanings that their components can have outside this unit (1992: 488). Expressions such as *csütörtököt mond* Thursday-ACC says 'it fails to work', *lépre csal* comb-SUB entices 'to toll' belong to this group.

Constructions where the verb has got a common noun argument belong to the second group. The meaning of the unit formed by the verb and its argument is not compositional, however, this construction cannot be an idiom for the noun preserves its original meaning, and, what is more, the basic meaning of the verb also plays an important role in computing the meaning of the construction. Expressions such as *moziba megy* lit. cinema-ILL goes 'to go to the cinema' or *iskolába jár* lit. school-ILL goes 'to go to school' compose this group.

The third group is formed by expressions such as *fejbe csap* lit. head-ILL hits 'to hit sy on the head' and *vállon csíp* lit. shoulder-SUP bites 'to bite sy on the shoulder'... Both components of the constructions are restricted semantically: the noun refers to a part of the body while the verb means physical contact.

The fourth group contains certain locutions. Their semantic head is the noun, the role of the verb is only to verbalize the construction. Examples of this type are *alkalom nyílik* lit. possibility opens 'a possibility emerges' and *módot ad* lit. way-ACC gives 'to provide an opportunity'...

Kiefer pays attention to constructions formed by a verb and a bare noun such as *újságot olvas* lit. newspaper-ACC reads '(s)he is reading a newspaper' (Kiefer 1990–91, 2003 and Kiefer and Ladányi 2000). On the basis of compositionality, he divides these constructions into two groups: productive constructions and idioms. According to his remarks, the bare noun before the verb is a verbal modifier and it is always an argument of the verb, on the other hand, it can never be a referential expression. A construction containing an accusative verb and a noun in the accusative case cannot be nominalised: the form **újságot olvasás* lit. newspaper-ACC reading 'reading a newspaper' is ungrammatical. However, the nominalisation of constructions with an intransitive verb is possible: in this case, it is the verb that is firstly nominalised, then the noun becomes its argument: (*moziba*) *jár – járás – moziba járás* lit. (cinema-ILL) goes – going – cinema-ILL going 'going to the cinema'. The formation of a present participle is similar to nominalisation, that is, the abovementioned restrictions hold for it, too. The complex verb formed by the noun and the verb becomes easily lexicalized, that is, the meaning of the construction becomes less transparent: in the constructions *ajánlatot tesz* lit. offer-ACC makes 'to make an offer', *vizsgát tesz* lit. exam-ACC makes 'to have an exam', *esküt tesz* lit. oath-ACC makes 'to make an oath' the same verb (*tesz*) can be found, however, the three actions are very different from each other.

2.2 A possible classification of bare noun + verb constructions

As it can be seen from earlier research, two factors, namely, compositionality and productivity play an important role when determining the relation between the verbal and the nominal component of bare noun + verb constructions (cf. Gábor and Héja (2006), and Kálmán (2006) on the relation of governors and their arguments). Based on these factors, in earlier work (Vincze 2008) I presented a bunch of grammaticality tests designed for the purpose to characterize bare noun + verb constructions more precisely. When designing different syntactic and semantic tests, I paid special attention to features of Hungarian bare common noun + verb constructions described earlier (Komlósy 1992, Kiefer 1990–91, Kiefer and Ladányi 2000 and Kiefer 2003) on the one hand, and to tests for English, German and French constructions given in Langer (2005) on the other hand. For the sake of illustration, I list some of the tests used: test of question (Can we ask a question for the nominal component?), test of nominalisation (Can the construction as a whole be nominalised?), omittance of the verb (In case of omitting the verb, can the original action be reconstructed?), and the test of variability (Can a single verb (derived from the stem of the noun) substitute the construction?) etc.

On the basis of the tests given in Vincze (2008), bare common noun + verb constructions can be divided into three groups. First, most of the tests give grammatical results for productive constructions. From the examples of Kiefer (1990–91) and Komlósy (1992), the following expressions are classified as productive constructions according to the results of the tests:

- (8) *újságot olvas* newspaper-ACC reads 'to read a newspaper' or *levelet ír* letter-ACC writes 'to write a letter' *moziba megy* lit. cinema-ILL goes 'to go to the cinema' or *iskolába jár* lit. school-ILL goes 'to go to school', *házat épít* lit. house- ACC builds 'to build a house'...

Constructions belonging to this group mostly describe conventionalized actions. Their structure is semantically transparent, their meaning can easily be calculated from the meaning of the verb, the noun and the suffix of the noun (that is, their compositionality is of high degree), therefore they are highly productive – this is why they are called productive constructions. (See Kálmán (2006) on the correlation between the degrees of compositionality and productivity).

Second, tests give ungrammatical results for idioms. Some examples can be seen in (9):

- (9) *csütörtököt mond* lit. Thursday-ACC says 'to fail to work', *gyökeret ver* lit. root-ACC beats 'to strike root'...

Constructions of this type are not semantically transparent, the meaning of the complex construction cannot be computed from the meanings of the parts of the expression, therefore their productivity is very low.

Third, there is a group of expressions for which some tests give grammatical, while other tests give ungrammatical results. This group of bare common noun + verb constructions will be hereafter called semi-compositional constructions because they

are situated in between the compositional productive constructions and the non-compositional idioms.

Semi-compositional constructions can be divided into subgroups on the basis of their behaviour being closer to the behaviour of productive constructions or the one of idioms. Constructions belonging to the first subgroup are situated closer to the productive constructions because they share more properties with productive constructions than idioms. This subgroup contains:

- (10) *előadást tart* lit. presentation-ACC holds ‘to give a lecture’, *parancsot ad* lit. order-ACC gives ‘to give an order’, *döntést hoz* lit. decision-ACC brings ‘to make a decision’, *bejelentést tesz* lit. announcement-ACC makes ‘to make an announcement’, *módot ad* lit. way-ACC gives ‘to provide an opportunity’...

Constructions of the second group are rather like idioms, concerning their behaviour. Such constructions are:

- (11) *virágba borul* lit. bloom-ILL falls ‘to come into bloom’, *igényt tart* lit. claim-ACC holds ‘to establish a claim’, *tetten ér* lit. act-SUP catch ‘to catch in the act’, *figyelembe vesz* lit. consideration-ILL takes ‘to take into consideration’, *áruba bocsát* lit. ware-ILL sends ‘to start to sell’, *lépre csal* lit. comb-SUB entices ‘to toll’ ...

Constructions of the third group are equally close to productive constructions and idioms, that is, they are not closer either to idioms or to productive constructions. Typical examples are:

- (12) *alkalom / lehetőség / esély nyílik / kínálkozik / adódik* ‘an opportunity / a chance / a possibility emerges’ ...

3 On the machine translatability of collocations

In machine translation, one of the most challenging tasks is the proper treatment of collocations. Every multiword expression is considered to be a collocation if its members quite often co-occur and its form is fixed to some extent (cf. Siepmann 2005, 2006, Sag et al. 2002, Oravecz et al. 2004, Váradi 2005). Examples are listed here:

- (13) *gyáva nyúl* lit. coward rabbit ‘chicken’, *hatos lottó* lit. six lottery ‘6/45 lottery’, *kreol bőrű* lit. creole skinned ‘dark-skinned’, *jóban-rosszban* lit. good-INE bad-INE ‘for better for worse’...

The translation of collocations is a hard task for both a human translator and a machine translation program, since their meaning is not totally compositional, that is, it cannot be computed on the basis of the meanings of the parts of the collocation and the way they are related to each other. Thus, the result of translating the parts of the

collocation can hardly be considered as the proper translation of the original expression.

3.1 Problems concerning the machine translation of collocations

When translating collocations, translation programs must face two main problems. On the one hand, parts of the collocation do not always occur next to each other in the sentence (split collocations). In this case, the computer must first recognize that the parts of the collocation form one unit (Oravecz et al. 2004), for which the multiword context of the given word must be considered. On the other hand, the lack (or lower degree) of compositionality blocks the possibility of word-by-word translation (Siepmann 2005, 2006).

In the following, some examples containing semi-compositional constructions are shown to illustrate the problems mentioned. The English sentences were translated by MetaMorpho, an English-Hungarian machine translation program (available at www.webforditas.hu). The results, that is, the Hungarian versions are given together with the precise English equivalent gained by retranslating the results into English. Finally, the correct Hungarian version of the original sentence is also presented.

- (14) All the trees have already come into bloom. (original English sentence)
Minden, ami a fáknak már van, bejön virágba. (MetaMorpho)
All that the trees already have enter into flower. (retranslation)
Már minden fa virágba borult. (Hungarian equivalent)
- (15) No lecture has he given this year. (original English sentence)
Nincs előadás neki ezt az évet adtak. (MetaMorpho)
There is no lecture to him this year they gave. (retranslation)
Ebben az évben még nem tartott előadást. (Hungarian equivalent)

In (14) there is a collocation (*come into bloom*) the parts of which occur next to each other, however, the program does not know this expression, that is why its parts are translated separately (word-by-word translation: *come* as *bejön* and *into bloom* as *virágba virág-ILL*). Since the expression is not compositional, the result is ungrammatical and meaningless. In (15) a split collocation can be found: parts of the collocation *give a lecture* are separated (other divergences from the word forms given by dictionaries are due to grammar). The program does not treat the collocation as a whole that is why no acceptable translation is provided.

3.2 A possible solution

Várad (2005) offers three different treatments of multiword expressions in computational linguistics. First, totally fixed expressions must be listed in the dictionary: the meanings of words in the English expression *French fries* do not equal to the ones of the words in the Hungarian equivalent *sült krumpli* lit. fried potato, thus, this expression must have a separate lexical entry in the dictionary. Second, productive expressions can be translated totally freely: in the case of *French wines*,

the translation of the parts of the expression provides the correct result (*francia borok*), thus, *French wines* does not form a separate lexical entry. Third, semi-fixed expressions are not worth listing in the dictionary because they are productive in the case of certain (semantic) groups of words. The scheme of the expression *French-speaking population* can be used for creating new expressions such as *Spanish-speaking population*, *Chinese-speaking population* etc. Local grammars have a leading role in the treatment of semi-fixed expressions in machine translation.

A parallel can be drawn between the three treatments of multi-word expressions and the three subgroups of bare noun + verb constructions. Productive constructions can be translated by using the word-by-word method, that is, they do not have to be listed in the dictionary, whereas idioms must be treated similarly to totally fixed expressions, thus, they must be listed in the dictionary. Nevertheless, semi-compositional constructions are too compositional for being listed in the dictionary since the relation between the parts of the constructions is constant. This relation can be formalized with the help of lexical functions, which will be presented in the following section.

4 Lexical functions and machine translation

4.1 Lexical functions

The theory of lexical functions was born within the framework of Meaning ↔ Text Theory (the model is described in detail in e.g. Mel'čuk 1974, 1989, 1996, 1998, Mel'čuk and Žolkovskij 1984, Mel'čuk et al. 1984–1999, Mel'čuk et al. 1995, Wanner 1997, 2007). The most important theoretical innovation of this model is the theory of lexical functions, which is universal: with the help of lexical functions, all relations between lexemes of a given language can be described. Although this theory has been thoroughly applied to different languages such as Russian, French, English or German, it has been rarely adapted to Hungarian: so far, it is only the applicability of **Magn** that has been studied (Répási and Székely 1998, Székely 2003).

Lexical functions have the form $\mathbf{f}(x) = y$, where \mathbf{f} is the lexical function itself, x stands for the argument of the function and y is the value of the function. The argument of the lexical function is a lexeme, while its value is another lexeme or a set of lexemes. A given lexical function always expresses the same semanto-syntactic relation, that is, the relation between an argument and the value of the lexical function is the same as the relation between another argument and value of the same lexical function. Thus, lexical functions express semantic relations between lexemes. In the case of syntagmatic lexical functions, these relations hold between expressions that are not totally compositional, that is, they must be learnt (Mel'čuk et al. 1995).

4.2 Lexical functions and semi-compositional constructions

4.2.1 Earlier research

Research on the relationship of semi-compositional constructions and lexical functions has been rarely conducted. One of the few exceptions is Apresjan (2004): he examines Russian verbal constructions that can be related to different lexical functions. He claims that there is a correlation between the given lexical function and, on the one hand, the type of the predicate, and, on the other hand, the semantic type of the nominal component of the construction. Since the meaning of the lexical function of **Oper** is given as follows: „delat' X, imet' X ili byt' v sostojanii X” 'to do X, to have X or to be in the state X'¹ (Apresjan 2004: 6), the values of **Oper**₁ will be such verbs whose meaning contains the element 'do' according to the definition.

Reuther (1996) focuses on three Russian support verbs – *vesti* 'comport', *provodit'* 'conduct', and *proizvodit'* 'manufacture'. He uses the Russian explanatory-combinatorial dictionary (Mel'čuk and Žolkovskij 1984) as a source of data. In his examples, the relation between the nominal component and the support verb can be formalized with the help of **Oper**₁. He concludes that nominal components can be divided into definite semantic groups in the case of all the three verbs: for instance, the support verb *provodit'* occurs together with nouns denoting an organized social activity or a complex procedure.

Studies on Russian material suggest that there is a correlation between the verb and the semantic type of the noun on the one hand, and between the lexical function and the type of the predicate on the other hand. Do these relations hold for Hungarian semi-compositional constructions as well?

4.2.2 Correlations between support verbs, nominal components and lexical functions

In earlier research on Hungarian support verb constructions (Vincze 2005), I found some relations holding between the verb, the semantic type of the nominal component and the lexical function used. Data were collected from the legal texts of the Szeged Corpus (for the detailed description of the corpus, see: http://www.inf.u-szeged.hu/projectdirs/hlt/index_en.html), a Hungarian database in which words are morphosyntactically analysed and tagged. I paid special attention to the support verbs *ad* 'give', *hoz* 'bring', *tesz* 'do', *vesz* 'take' and *végrehajt* 'realise'. The relations between the support verbs and the nouns are formalised in terms of lexical functions, that is, it is revealed which support verb is the value of which lexical function in the case of a specific noun. Thus, correlations between support verbs and semantic classes of nouns can be formulated: namely, special semantic classes co-occur with special support verbs. For instance, nouns denoting speech acts tend to co-occur with the support verbs *ad* 'give', *tesz* 'do' or *hoz* 'bring', while nouns denoting a possibility usually co-occur with *ad* 'give'.

The meaning of support verbs found in dictionaries and the semantic content of lexical functions were also compared. It was revealed that their semantic components must (partly) overlap. For instance, support verbs whose meaning contains the

¹ The translation is mine (VV).

semantic primitive ‘do’ (*ad* and *tesz* in the database) are usually values of the lexical function **Oper**₁. The verb *vesz* contains the semantic element ‘begin’, that is why it is usually the value of the lexical function **Incep** referring to the beginning of an event. Both the support verb *végrehajt* and the lexical function **Real** contain the semantic element ‘fulfil the expectations’, and the common part in the meaning of *hoz* and **Caus** is the semantic primitive ‘cause’.

The results of this study suggest that there are correlations between the verb and the semantic type of the noun, and specific verbs are often values of specific lexical functions in the Hungarian language as well.

4.2.3 Lexical functions and subgroups of semi-compositional constructions

Let us return to the grouping of semi-compositional constructions. As it was presented earlier (see 2.2), these constructions can be divided into three groups. Different groups can be related to different lexical functions. First, in the case of constructions being closer to productive constructions, the nominal complement tends to be the syntactic object of the verb, for instance: *előadást tart* lit. presentation-ACC holds ‘to give a lecture’, *parancsot ad* lit. order-ACC gives ‘to give an order’, *döntést hoz* lit. decision-ACC brings ‘to make a decision’... In the language of lexical functions, this syntactic relation can be expressed by **Oper** (Mel’čuk et al. 1995), as in:

(16) **Oper**₁ (*döntés*) = [*~t*] *hoz* ‘make a decision’.

Second, in the case of constructions being closer to idioms, the nominal component usually has an oblique case: *virágba borul* lit. bloom-ILL falls ‘to come into bloom’, *számításba vesz* lit. account-ILL takes ‘to take into account’, *figyelembe vesz* lit. consideration-ILL takes ‘to take into consideration’... This syntactic relation is described by the lexical function **Labor** (Mel’čuk et al. 1995):

(17) **Labor**₁₂ (*számítás*) = [*~ba*] *vesz* ‘take into account’.

Third, the nominal component of construction being in the middle, that is, in between productive constructions and idioms functions as a subject: *alkalom nyílik*, *esély kínálkozik*, *lehetőség adódik* ‘an opportunity / a chance / a possibility emerges’ ... This syntactic relation can be expressed by the lexical function **Func** (Mel’čuk et al. 1995):

(18) **Func**₁ (*alkalom*) = *nyílik* ‘a possibility emerges’.

Now it is highlighted that the subgroups of semi-compositional constructions correlate to the groups of lexical functions, more precisely, a specific subgroup can be paired with a specific lexical function. Since it has been already stated (in 4.2.2) that there is a correlation between lexical functions and the semantic content of their values (that is, of the verbs), it can be expected that certain groups of semi-compositional constructions often contain specific verbs. Thus, constructions being closer to idioms contain verbs that do not typically occur in constructions being closer to productive constructions and vice versa.

4.3 Semi-compositional constructions, lexical functions and machine translation

The results of my earlier studies on Hungarian data (Vincze 2005, 2008) suggest that on the one hand, the semantic type of the noun can predict what verb will be the value of a given lexical function, being the noun its argument, and, that on the other hand, certain verbs are typical values of a given lexical function. Furthermore, in this paper I have shown that the groups of semi-compositional constructions tend to correlate with the groups of lexical functions. Thus, in Hungarian, it can be predicted to a certain degree what verb will co-occur with a given noun in the case of a given syntactic relation. Here I provide an example: if the accusative form of the noun *tájékoztatás* 'information', that is, *tájékoztatást* needs a support verb, then two facts must be considered. First, the accusative form of the noun refers to the predicate-object syntactic relation, which is described by **Oper**, and verbs such as *ad* 'give', *tesz* 'make', *hoz* 'bring', *vesz* 'take', *kap* 'get' are the most common values of this lexical function in Hungarian. Second, *tájékoztatás* is a noun denoting a verbal act, and nouns denoting verbal acts or speech acts are usually paired with verbs like *tesz*, *ad* and *hoz*. Thus, the constructions *tájékoztatást ad* and *tájékoztatást tesz* can be predicted (actually, both constructions can be found in the legal texts of the Szeged Corpus).

These results can be fruitfully applied in machine translation as well. However, for a successful translation, these generalizations must be made for both the source language and the target language. Invaluable sources of these generalizations are explanatory combinatorial dictionaries (Mel'čuk and Žolkovskij 1984 for Russian and Mel'čuk et al. 1984–1999 for French) containing different semantic and syntactic relations between lexemes coded by the means of lexical functions.

The applicability of lexical functions in machine translation is emphasized in Apresjan and Cinman (2002). The two languages they focus on are English and Russian. If a list containing all the values of lexical functions applied to a lexeme is available for both languages, machine translation becomes much easier and more precise, since it is only the two lists that must be compared and the corresponding lexeme can be easily chosen (as an illustration, I provide the Hungarian equivalent of this construction as well):

(19) **Oper**₁ (*nadežda*) = [*~u*] *pitat*'

(20) **Oper**₁ (*hope*) = *cherish*

(21) **Oper**₁ (*remény*) = [*~t*] *táplál*

To sum up, the machine translation of semi-compositional constructions can be supported in two ways. First, constructions can be stored in the form of lexical functions: in this case, the size of the dictionary grows but the translations become more precise. Second, a translation can be proposed on the basis of the connection between the semantic content of the noun and the verb. The latter is a statistical-based method: for each noun belonging to a certain semantic group, it is highly probable that it is accompanied by a certain verb, for instance, nouns denoting speech acts usually for a semi-compositional construction together with one of the verbs *ad* 'give', *tesz* 'make' and *hoz* 'bring' in Hungarian (Vincze 2005), and with the verbs *davat*' or *delat*' in Russian (Apresjan 2004). The machine translation program

chooses the appropriate verb for each noun with the help of frequency rates based on data in a training corpus: in this way, when translating the semicompositional construction *eskiüt tesz* 'make an oath' to Russian, the noun *kljatva* will be paired with the verb *davat* 'since *davat* *kljatvu* occurs much more frequently than *?delat* *kljatvu*. However, both methods need thorough preparation: either the lists containing the values of lexical functions must be created for both languages, or the semantic connections between nouns and verbs in both languages must be described in detail. Since these tasks have not been adequately performed for Hungarian (my research being only a first step in this direction), long theoretical and empirical work is still needed.

5 Conclusion

In this paper, I have presented some possible ways of translating semi-compositional constructions with the help of a computer. Two possible methods have been proposed: first, the one based on lexical functions, and, second, the statistical-based one making use of lexico-semantic relations between the noun and the verb. The application of both methods in machine translation would yield a much more precise result. However, both methods require previous theoretical research, the results of which could be fruitfully applied in the field of computational linguistics, especially in machine translation. Hopefully, this theoretical work will be performed in the near future.

References

- Apresjan, Ju. D. 2004. O semantičeskoj nepustote i motivirovannosti glagol'nyx leksičekix funkcij. *Voprocjazykoznanija* Vol. 4. 3-18.
- Apresjan, Ju. D., Cinman, L. L. 2002. Formal'naja model' perifrazirovanija predloženiij dlja sistem pererabotki tekčkov na estestvennyx jazykax. *Russkij jazyk b naučnom osveščeniij* Vol. 2. No. 4. 102-146.
- Dobos, Cs. 1991. *Leíró kifejezések az orosz jogi szaknyelvben [Descriptive expressions in Russian legal language]*. Doctoral dissertation, ms.
- Dobos, Cs. 2001. *A funkcióigés szerkezetek vizsgálata (különös tekintettel az orosz jogi szaknyelvre) [A study on support verb constructions – focusing on Russian legal language]*. PhD dissertation, ms., Debrecen.
- Gábor, K., Héja, E. 2006. Predikátumok és szabad határozók [Predicates and free circumstantials]. In: Kálmán, L. (ed.): *KB 120 – A titkos kötet. Nyelvészeti tanulmányok Bánréti Zoltán és Komlósy András tiszteletére*. Budapest: Tinta Kiadó. 135-152.
- Grétsy, L., Kemény, G. 1996. *Nyelvművelő kéziszótár [Purists' dictionary]*. Budapest: Auktor Könyvkiadó.
- Heltai, P., Gósy, M. 2005. A terpeszkedő szerkezetek hatása a feldolgozásra [The effect of sprawling constructions on processing]. *Magyar Nyelvőr* Vol. 129. No. 4. 471-487.
- Kálmán, L. 2006. Miért nem vonzanak a régensek? [Why do governors not have arguments?]. In: Kálmán, L. (ed.) *KB 120 – A titkos kötet. Nyelvészeti tanulmányok Bánréti Zoltán és Komlósy András tiszteletére*. Budapest: Tinta Kiadó. 229-246.

- Keszler, B. 1992. A mai magyar nyelv szófaji rendszere [The system of parts-of-speech in contemporary Hungarian]. In: Kozocsa, S. G., Laczkó, K. (eds.) *Emlékkönyv Rácz Endre hetvenedik születésnapjára*. Budapest: ELTE. 131-139.
- Kiefer, F. 1990–91. Noun Incorporation in Hungarian. *Acta Linguistica Hungarica* Vol. 40. No. 1-2. 149-177.
- Kiefer, F. 2003. A kétféle igemódosítóról [On the two types of verbal modifiers]. *Nyelvtudományi Közlemények* Vol. 100. 177-186.
- Kiefer, F., Ladányi, M. 2000. Az igekötők [Verbal particles]. In: Kiefer, F. (ed.) *Strukturális magyar nyelvtan. 3. Alaktan*. Budapest: Akadémiai Kiadó. 453-518.
- Komlósy, A. 1992. Régensék és vonzatok [Governors and arguments]. In: Kiefer, F. (ed.) *Strukturális magyar nyelvtan I: Mondattan*. Budapest: Akadémiai Kiadó. 299-527.
- Langer, S. 2005. A Linguistic Test Battery for Support Verb Constructions. *Linguisticae Investigationes* Vol 27. No. 2. 171-184.
- Mel'čuk, I. 1974. Esquisse d'un modèle linguistique du type "Sens<>Texte". In: *Problèmes actuels en psycholinguistique*. Colloques inter. du CNRS, n° 206. Paris. 291-317.
- Mel'čuk, I. 1989. Semantic Primitives from the Viewpoint of the Meaning-Text Linguistic Theory. *Quaderni di Semantica* Vol. 10. No. 1. 65-102.
- Mel'čuk, I. 1996. Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon. In: Wanner, L. (ed.) *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/Philadelphia: Benjamins. 37-102.
- Mel'čuk, I. 1998. Collocations and Lexical Functions. In: Cowie, A. P. (ed.) *Phraseology. Theory, Analysis, and Applications*. Oxford: Clarendon Press. 23-53.
- Mel'čuk, I. et al. 1984–1999. *Dictionnaire explicatif et combinatoire du français contemporain: Recherches lexico-sémantiques I–IV*. Montréal: Presses de l'Université de Montréal. 1984, 1988, 1992, 1999.
- Mel'čuk, I., Clas A., Polguère A. 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.
- Mel'čuk, I., Žolkovskij, A. 1984. *Explanatory Combinatorial Dictionary of Modern Russian*. Vienna: Wiener Slawistischer Almanach.
- Oravecz, Cs., Varasdi K., Nagy V. 2004. Többszavas kifejezések számítógépes kezelése [The treatment of multiword expressions in computational linguistics]. In: Alexin, Z., Csendes, D. (eds.) *MSzNy 2004 – II. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem. 141-154.
- Répási, Gy., Székely, G. 1998. Lexikográfiai előtanulmány a fokozó értelmű szavak és szókapcsolatok szótárához [A lexicographic pilot study on the dictionary of intensifying words and collocations]. *Modern Nyelvoktatás* Vol. 4. No. 2-3. 89-95.
- Reuther, T. 1996. On Dictionary Entries for Support Verbs: The Cases of Russian VESTI, PROVODIT' and PROIZVODIT'. In: Wanner, L. (ed.) *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/Philadelphia: Benjamins. 181-208.
- Sag, I. A., Baldwin T., Bond F., Copestake A., Flickinger D. 2002. Multiword Expressions: A Pain in the Neck for NLP. In: Gelbukh, A. (ed.) *Proceedings of CICLING-2002*. Mexico City.
- Siepmann, D. 2005. Collocation, colligation and encoding dictionaries. Part I: Lexicological Aspects. *International Journal of Lexicography* Vol. 18. No. 4. 409-444.
- Siepmann, D. 2006. Collocation, colligation and encoding dictionaries. Part II: Lexicographical Aspects. *International Journal of Lexicography* Vol. 19. No. 1. 1-39.
- Székely, G. 2003. *A fokozó értelmű szókapcsolatok magyar és német szótára [Hungarian – German dictionary of intensifying collocations]*. Budapest: Tinta Könyvkiadó.
- Szikli, L. 1986. Terpeszkednek vagy körülírnak? [Do they sprawl or do they circumscribe?] *Magyar Nyelvőr* Vol. 110. No. 3. 268-273.
- Váradí, T. 2005. Többszavas kifejezések kezelése MT szótárban [The treatment of multiword expressions in MT dictionaries]. In: Alexin, Z., Csendes, D. (eds.) *MSzNy 2005 – III*.

- Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem. 233-244.
- Vincze, V. 2005. *Funkcióigés szerkezetek vizsgálata lexikai függvények segítségével [An analysis of Hungarian support verb constructions with the help of lexical functions]*. Presented at: 9th National Conference of PhD Students in Linguistics, Szeged. 18th November 2005.
- Vincze, V. 2008. A puszta köznév + ige komplexumok státusáról [On the status of bare common noun + verb constructions]. *LingDok 7. Nyelvész-doktoranduszok dolgozatai*. Szeged: Szegedi Tudományegyetem. 265-283.
- Wanner, L. (ed.) 1997. *Recent Trends in Meaning-Text Theory*. Amsterdam/Philadelphia: Benjamins.
- Wanner, L. (ed.) 2007. *Selected Lexical and Grammatical Issues in the Meaning-Text Theory. In honour of Igor Mel'čuk*. Amsterdam/Philadelphia: Benjamins.