COMPLEMENTS AND ADJUNCTS IN MACHINE TRANSLATION

Kata Gàbor, Enikö Héja
Eötvös Loránd University and Hungarian Academy of Sciences

Abstract

A significant number of natural language processing applications cannot work without syntactic parsing. The automatic syntactic analysis of natural language texts in turn requires an efficient method for differentiating between elements that belong to the predicate's argument structure and those that are attached to it as adjuncts. The focus of our paper is a specific method we are working on for differentiating between verbal complements and adjuncts, which we intend to use for the elaboration of a Hungarian verbal argument structure database, particularly suited for machine translation purposes.

1. Introduction

Both linguistic theories and rule-based natural language processing applications rely on a strict differentiation between verbal complements (elements that figure in the subcategorization frame of the verb's lexical entry) and adjuncts (elements that are optionally added to the verb (phrase) by syntactic rules). Although adjuncts are optional and hence their appearance is not predictable, the possibility to extend a verbal structure by an adjunct is predictable. In opposition, the behavior of complements is not predictable by general syntactic rules of a given language, this is why they are widely conceived as lexical properties of verbs and they are treated in the lexicon. Consequently, Natural Language Processing (NLP) applications that involve syntactic parsing of texts need to use a lexical database of verbal argument structures which describe all the relevant properties of every single verb's arguments. However, for the database to be coherent and homogeneous, coders need to be given exact and explicit instructions about what a complement is. This boils down to our basic question: what could be the method for making the difference between complements and adjuncts?

We examined two linguistic theories: Government and Binding theory (GB; on the basis of Radford 1988) and Lexical Functional Grammar (LFG; on the basis of Komlósy 2001) with respect to how they describe verbal argument structure and the way arguments are represented in the surface structure of natural languages. The most significant difference between these theories is that GB proposes a configurational model of natural languages, i.e. it encodes constituents' grammatical functions

[ASJU, XLI-2, 2007, 59-70]

by dominance and precedence relations in the tree structure, while in LFG grammatical functions are coded in a separate level of representation which does not prescribe their possible surface representation. The importance of this difference is that in several languages complementness and syntactic functions are reflected not as much in surface constituent order as in morphological properties. In Hungarian configurationality is used to express discourse functions instead of syntactic functions, thus we would predict that an LFG analysis would describe better complementness criteria in Hungarian. On the other hand, both theories agree that complementness is a relational notion: a given element can only be the complement of a governing element, but not in itself.

2. The role of complements in machine translation systems

Among current theoretical and methodological approaches to machine translation (MT) two main branches can be distinguished (Jurafsky and Martin 2000): *rule-based* and *statistical/corpus-based* systems. Rule-based systems use linguistic knowledge: they contain one or more modules which analyze source language text units at several linguistic levels, and rules map the output of the source language analysis to the target language or to an intermediate representation.

The main advantage of rule-based MT systems as opposed to statistical ones is that they are more easily maintainable: to find the source of an incorrect translation is relatively trivial in a well-designed rule-based system, while it can be very complicated in a statistical one.

Rule-based MT systems can be subdivided into three types:

1) *direct transfer* 2) *transfer-based translation* 3) *interlingual translation*

As its name implies, direct transfer is a simple method based on the supposed similarity between closely related languages: it makes wide use of bilingual dictionaries, but does not direct much attention to structural differences. The role of grammatical rules for source language analysis and translation is marginal: they mainly serve as disambiguation rules.

Transfer-based systems analyze the source language text at both morphological and syntactic levels by monolingual rules and databases, and use the so-called *transfer rules* to map the output of the analysis into the target language. The final stage of the translation is the set of monolingual target language grammar rules which correct the output of the transfer phase. The key module of the process is the bilingual *transfer module* which is composed of the bilingual dictionary and the transfer rules. These rules carry out the task of mapping grammatical characteristics of the source language into the target language. Hence, this module is totally specific to language pairs as it only deals with phenomena that differ across the given language pair.

As opposed to transfer-based translation, *interlingual* systems project the source language text into an intermediate representation which is a language-independent structure intended for outlining the information contained in the sentence as well as its logical structure. Target language equivalents of the sentence are then calcu-

lated from the intermediate representation. While transfer-based systems attain the target language translation of a sentence by means of transforming the elements and the structure of the original source language text, interlingual systems aim at extracting the meaning of the source text and produce a target language text with the same meaning.

Both transfer-based and interlingual MT systems rely strongly on the syntactic parsing of the source language, all the more because it plays an important role in disambiguation. Most systems lay emphasis on setting apart lexical information and general sentence formation rules. The reason for it is the assumption that while the translation of lexical information is unpredictable, regular phenomena can be translated by rules to another natural language (in transfer-based systems) or to an intermediate representation (in interlingual systems). This distinction applies to verbal subcategorization and adjunction. On the one hand, syntactic behavior of the elements which fulfill complement or adjunct functions can be predicted within a given language and translated by rules: e. g. we can state that in Hungarian the top-level NP constituent in nominative case will be the subject of the clause, and construct a rule which translates it into English by moving this NP before the verb. On the other hand, whether a given Hungarian verb *can have* a subject and whether it will keep this function throughout translation is a piece of unpredictable, though important information which has to be coded.

3. Tests for complementness

We made a comparison between a configurational and a lexicalist linguistic theory (Government and Binding theory and Lexical Functional Grammar, respectively) with respect to how they describe verbal argument structure and how they represent arguments in the surface structure of natural languages.

3.1. Complements and subcategorization in GB

Government and Binding theory defines complements as constituents which compulsorily appear in the close local context of the verb. Their syntactic behavior cannot be described by general phrase-structure rules as their appearance is not predictable. The reason for this is that complementness is conceived as a relation: constituents which fulfill a complement function in a sentence with respect to its predicate cannot have this same function in other sentences. Predicates' ability to take complements is their idiosyncratic lexical property. Consequently, lexical entries of verbs have to contain as much information as necessary for the syntactic rules to generate surface forms of complements. Hence, lexical entries of verbs comprise syntactic description (i.e. the category) of their complements. Moreover, since syntactic complements are surface representations of semantic arguments, it is worth coding the thematic roles of semantic arguments in the lexicon as it allows certain generalizations over the surface representation of semantic arguments: a part of the complement structure can be derived from thematic roles.

3.2. Complement tests in GB

According to X-bar theory, if we want to test whether a given constituent is a complement or an adjunct we have to examine its structural position. While complements are located within the syntactic tree in a sister node of the X (verbal) head and together they form an X' projection, adjuncts are sisters of the X' projection and form a new X' with it. The position that adjuncts and complements occupy in the syntactic tree is universal among languages, but their surface order in relation to the head is language-specific. Unfortunately, this implies that we can only rely on language-specific tests for verifying the different structural position of given constituents. Radford (1988) mentions four tests for English:

a) *Passivization*: NPs raised from a complement PP can be passivised while NPs from an adjunct PP cannot:

[This job] needs to be worked at by an expert.
*[This office] is worked at by a lot of people.

b) *Pronominalization*: the *do so* structure, which replaces the category V', can include adjuncts that are attached to a V' to form a new V' projection with it (i), but adjuncts can also be omitted from it (ii), while complements are compulsorily included (iii), they cannot be omitted as in (iv).

i) John will [buy the book on Tuesday] and Paul will *do so* as well.
ii) John will [buy the book] on Tuesday and Paul will *do so* on Thursday.
iii) John will [put the book on the table] and Paul will *do so* as well.
iv) *John will [put the book] on the table and Paul will *do so* on the chair.

c) *Surface order*: Complements are closer to the verb than adjuncts because they connect to the verb in the syntactic tree earlier than adjuncts, and crossing branches are forbidden.

d) *Ellipsis*: Any phrasal category can be subject to ellipsis. Constituents of the category of V' can be ellipsed if they consist of the verbal head with its complements and adjuncts (i), the head with its complements but without adjuncts (ii), but the head with one of its complements and without the other one does not form a constituent, hence it cannot be ellipsed (iii):

i) — Who might be going to the cinema on Tuesday?
— *John might be …*
ii) — Who might be going to the cinema when?
— *John might be … on Tuesday.*
iii) — Who will put the book where?
iv) **John will … on the table.*

3.3. Complements and subcategorization in LFG

In the LFG model different structural levels of the sentence carry the same functional information but are represented at different levels. However, information about grammatical function is present at every level of representation. Accordingly, grammatical information is represented at three levels:

1. a-structure stores lexical information, i. e. argument structure;
2. c-structure stores surface constituent structure;
3. f-structure represents the language-independent functional structure which can be extracted from the two language-specific representations.

Surface structure is not an independent level of representation: it is generated from c-structure by inserting lexical elements.

Argument structure and other lexical information are stored in a-structure. Arguments are conceived as 'unfilled' slots in the meaning of the verb. Accordingly, the bare a-structure comprises semantic arguments of the verb with their thematic roles. Surface representation of semantic arguments depends on the grammatical functions associated with them. The first step in the process of mapping argument structure to surface complements is the annotation of bare lexical structure. Annotation assigns a function to semantic arguments. In LFG verbal subcategorization frames do not contain categorial information about complements, they only refer to their grammatical function. Correspondences between semantic argument positions and grammatical functions are coded in the annotated lexical structure —argument roles cannot be bound to universal structural positions (as in GB, where subjects and direct objects are assigned their function on the basis of their structural positions).

The set of functions verbs can prescribe for their complements is restricted. There are complement and adjunct functions. Among complement functions, the most interesting distinction is between *thematically bound oblique complements* and *labeled complements*. Thematically bound oblique complements are those complements whose thematic role is determined by the verb, but whose syntactic functor is not. One typical example is constituted by locative complements. On the other hand, in the case of labeled complements, not only the thematic role but also the exact form of the syntactic functor (e.g. its case suffix or preposition) is prescribed by the verb.

Surface representation of constituents with complement functions is generated by language-specific context free rules of the c-structure. Surface representation of grammatical functions may be coded either configurationally or by other (typically morphological) means, and this divergence may be present within one language. In the case of configurationally coded functions, the LFG variant of X-bar theory conditions the construction of the tree structure, while in the case of non-configurational coding, complement functions are associated not to the c-structure but to case-marking and agreement properties. The way functions are associated to case-bearing constituents is by functional annotation schemes realized as implications: "*a constituent X may be associated with a function Y if it bears case Z*." On the other hand, agreement is handled by *head marking*: the constituent's relevant AGR features are checked, and the constituent is associated with the given function if the value of the features equals those prescribed by the functional structure of the constituent which dominates it.

3.4. Complement tests in LFG

Komlósy (1992), in his LFG analysis on Hungarian verbal argument structure, defines complements as elements whose syntactic and semantic properties are subcategorized by the governing verb. For analyzing given structures, he suggests using the following three tests:

e) if a constituent is obligatory in any level of sentence structure, it is a complement;
f) if a constituent's appearance in the structure allows to expand it further by an optional adjunct (which could not be present in the original structure), then this constituent is an (optional) complement;
g) if a word X has an expansion Y, and there is a word Z which can systematically replace X+Y, and can replace X when X is not expanded by Y, but cannot replace X when Y is present, then Y is an optional complement of X.

3.5. Hungarian syntax

Hungarian is a highly inflective language with 18 cases and a (roughly) free word order: this means that almost any ordering of the verb and its expansions is acceptable, although they yield slightly different interpretations. As described in É. Kiss (2002), in the neutral sentence verbal complements and adjuncts follow the predicate within the VP. However, in most sentences, at least one verbal complement precedes the verb —this is the *topicalized* constituent, which is raised to the first position in the sentence. Another syntactic movement that changes the neutral word order is *focusing*: the focus position is the one immediately preceding the finite verb. Any type of complements or adjuncts can be topicalized and focused, hence moved outside the VP. Furthermore, *verb modifiers,* i. e. verbal prefixes, adverbs or bare NP complements also precede the verb they modify. Thus, most verbal complements and adjuncts are free to appear before the predicate. When parsing Hungarian texts, we face the difficulty of being unable to determine dependency relations and grammatical functions on the basis of constituent order. On the other hand, Hungarian morphology is very rich, thus we have to rely on constituents' morphological features, in particular on case marking. In compliance with these features of Hungarian, we find that most of the above-mentioned GB and LFG-related tests either do not apply or are insufficient for deciding on complementness.

GB tests

Tests a) and b) do not apply to Hungarian due to the lack of passivization and pronominalization. As to condition c), it is not always met in the surface order of Hungarian sentences:

A gyerekek nyírják a kertben a füvet.
The children[NOM] cut[PL.3] the garden[INE] the grass[ACC].
The children are cutting the grass in the garden.

Bea megtalálta tegnap a kutyádat.
Bea found yesterday the dog[POSS.S2+ACC]
Bea found your dog yesterday.

In the sentences above, adjuncts (a locative NP in inessive case: *a kertben*-'in the garden' and a temporal adverb: *tegnap*, 'yesterday') precede the obligatory complements (direct objects in accusative case: *a füvet* —'the grass' and *a kutyádat*— 'your dog') and hence they are wedged in between the verb and its complements.

Assuming that *tesz*, the Hungarian counterpart of 'put', similarly to the English verb has three argument places, we can prove that condition d) concerning ellipsis possibilities within the VP does not hold for Hungarian:

Ki megy hová kedden?	— János … moziba …
Who goes where on Tuesday?	— John … to the movie …
Ki tette a könyvet hová?	— János … *az asztalra.*
Who put the book where?	— John … on the table.

LFG tests

There is a counter-argument for e) (which only holds for obligatory complements): any of the verbal complements can be omitted in Hungarian. As for f) and g) tests, they can only be applied to a limited number of verbs: the expansion of the structure by a complement does not always entail the possible appearance of an adjunct (as in f), and we cannot be sure to find a synonym (with different valence) for each verb we are dealing with (as in g). Thus these criteria do not seem to suffice for our purposes.

4. Compositionality as a criterion

Lexical entries of verbs thus contain those elements that appear in the local context of the verb and which cannot be derived by general phrase structure rules. In compliance with GB, we would like to construct lexical entries whose subcategorization frames specify the category and (in the case of postpositional complements) the lemma of the complements. On the other hand, instead of relying on the local context, we loosen this constraint and look for complements and adjuncts in the whole extent of the clause which contains the finite verb. The reason for it is the phenomenon known as "scrambling": certain non-configurational languages with a rich morphology show a much bigger diversity in the surface order of sentence constituents than, for instance, English. Complements and adjuncts are free to mingle, they might even leave their clause after having received a case from the verb.

This implies that surface order cannot be used for separating complement and adjunct functions. On the other hand, the markedly rich morphological system can serve as a basis for our investigations. Thus, instead of using configurational information, we intend to use morphology, especially case feature as a marker of the syntactic role. The basic assumption is that not every occurrence of an NP with a case suffix is lexically subcategorized by a verb: some of them are added to the sentence by productive rules. Since such rules prescribe as a syntactic requirement the appearance of a certain case suffix on the NP and associate a syntactic role to it, we conclude that these suffixes are elements which enable NPs to fulfill certain roles. In other words, default meanings can be associated with case suffixes. Another important assumption we relied on is that once we manage to state the function of the NP with a particular case, its translation can be generated from the translation of the NP by the application of translation rules. Hence our definition of complementness and adjunctness will be based on their degree of compositionality.

With these presuppositions, our work starts by enumerating possible syntactic and semantic functions of case suffixes. This means that we try not to see complements in their relation to the predicates, and we conceive the predicate-argument functions as one possible function that case suffixes may bear. We find that there are two grammatical cases (nominative and accusative) that cannot have a default meaning and can only occur with verbal complements. These cases have to be included in verbal valence structures. As for the other cases, we try to define all their syntactic and semantic properties that can be described —and translated in a machine translation system— by general rules. Such rules specify one or more translations for the given case suffix, and may refer to semantic or syntactic features of the constituent they appear in, but general rules may not refer to the predicate. For example, the case suffix *-ban* ("inessive case") indicates the exact date if it appears on a constituent expressing time: it forms a regular adjunct of time. Otherwise, it expresses location, and also forms a regular adjunct. These two rules will work as *default rules* for the case suffix *-ban*, assuming that in absence of lexical rules, the case is associated to one of these functions independently of the context:

[NP.case = ins,semantics=time] → [NP.role = time.adjunct]
[NP.case = ins] → [NP.role = loc.adjunct]

The default rule is conceived as a definition of the relation between the case-bearing element and the verb. In the example above, NPs get labeled as time or locative adjuncts which specify the kind of information they add to the predicate's meaning.

While defining default rules, it comes into light that some [V + NP.case] structures are midway between rule-based constructions and total lexicalism. This means that their function can be stated, but their appearance depends on the semantic class of the predicate. For example, the ablative case *-tól* may have two default meanings: with movement verbs it marks the starting point of a movement; with verbs that express a change in someone's state, it expresses the cause of the change. Thus, we can associate a function to the case suffix with rules that refer to the semantic class of the verb it occurs with. This kind of rule cannot be considered as default because it refers to the predicate. However, we find that even the appearance of regular adjuncts like time adverbs are constrained by the semantics of the verb they modify, but still we would not like to consider them as being part of the verb's valence. These semi-productive rules represent a new category in-between complements and adjuncts: accordingly, when performing syntactic parsing, their application follows verbal valence matching but precedes default rules.

Because semi-default rules refer to verb classes when they apply, we had to tackle the task of creating predicate classes on reasonable grounds. The main characteristics of the classes are described by metapredicates. At this point we assume that systematic *syntactic* and *morphological* alternations are able to serve as good hints while defining our metapredicates. This presupposition is in accordance with what we stated before, namely, the NPs' syntactic function in relation to the predicate of the clause depends on the existence of the case suffix's meaning. This meaning also determines the given NP's semantic relation to the verb.

First, let us examine the criterion of morphological changes of the predicate. Productive derivational processes change the meaning of the verb in a systematic way,

and sometimes also the argument structure of it. When the argument structure does not change we can conclude that the meaning denoted by the derivational suffix is of no importance regarding the NPs' syntactic or semantic role in the sentence, or that the given NP might be a complement the case suffix of which bears no meaning at all. As for the first possibility, the interesting thing here is that in most cases we cannot make such generalizations over the totality of verbs. This means that usually even if a derivational suffix does not have an effect on the argument structure of the majority of predicates, there might be also verbs whose argument structure does undergo certain changes. Indeed, this difference strongly implies the metapredicates we should use when describing a verb class, since we might expect that when the NP with its case suffix stays as it was, the meaning of the case suffix is compatible with that of the derivational suffix, while regarding the other case the meanings are controversial that is why the argument structure changes along with the meaning of the predicate itself.

For instance, *-gAt* is a derivational suffix which expresses two different *aktionsarts*, diminutive and iterative —depending on some semantic features of the base verb (Kiefer, Ladányi 2000). Usually *-gAt* does not change the argument structure of the base verb. Just as in the following verb pairs:

olvas - olvasgat 'keep reading'; *lövöldöz vkre - lövöldözget vkre* 'keep shooting at sg'

In fact, *-gAt* may also attach to verbs whose argument structure do change:

iszik vmre 'drink to sg' vs. **Iszogat vmre* 'keep drinking to sg'*;*

halaszt vmit vmire 'postpone sg to sg' vs **halogat/halasztgat vmit vmire* 'keep postponing sg to sg';

From the examples above it follows that there is a verb class which has a meaning component —expressed by the relevant NP's case suffix— that is not compatible with the meaning of *-gAt.*

Another basis we use to distinguish among predicate classes is the systematic change in the argument structure. It could be stated that from a linguistic point of view this criterion is the same as the morphological one. To supply proof for this hypothesis we only need to stipulate the existence of some zero-morphemes which are responsible for the alternation of argument structures. On the one hand this method is supposed to guarantee the reliability of the metapredicates. On the other hand we expect that these alternations have no effect on complements but on adjuncts, i.e. complements also might appear in the structure after the derivation with the same suffix. This follows from our presupposition, namely that in the case of complements the case suffix appearing on the head of the NP adds nothing to the whole structure's meaning. While derivational suffixes are considered as forms that yield always the same meaning —and obviously that is the way we want to look at them— we can say that derivational suffixes are functions that take only the verbs' meaning as input. As opposed to adjuncts, complements' suffixes play no role from the point of view of the verb's meaning, hence we would expect that any change of the verbs' meaning would leave such suffixes untouched. Regarding adjuncts we expect that —because the case suffix meaning is composed with the predicate meaning— they can undergo alteration when the appropriate derivational suffix is attached to the base verb.

Now we present how we use the conditions above in the case of the Hungarian suffix *-tÓl* (instrumental case). What kind of restrictions can be formulated regarding the three different argument structures below?

[11.a] *János felébresztette Mari - t a zaj - jal.*
John awoke Mary - ACC the noise - INS
'John awoke Mary with the noise.'

[11.b] *A zaj felébresztette Mari - t.*
The noise awoke Mary - ACC
'The noise awoke Mary.'

[11.c.] *Mari felébredt a zaj - tól.*
Mary woke up the noise - ABL
'Mary was awoken by the noise.'

We supposed that the semantic representation of the verbs belonging to this class is as follows:

CAUSE (John, E), where E < noise, CHANGE(S, S') > and CAUSE(cause, S')

which means that John brought (CAUSE) a situation (E) into existence, and E is a two-argument predicate, such that there is an x ('noise'), which causes (CAUSE) a change in Mary's *mental* state, namely a change from S into S'. The next question is, how could we verify syntactically these three semantic components (i.e. CAUSE, MENTAL, CHANGE)? We suppose that a verb belongs to this class if and only if it can undergo systematically the syntactic alternations represented in [11.a.], [11.b.] and [11.c.].

As [11.a.] and [11.b.] show, the predicates belonging to this verb class have to have at least one interpretation where the subject is non-agentive. Otherwise [11.b.] should be ungrammatical, since the denotata of such subjects cannot carry out an action voluntarily. This requirement is responsible for the fact that most verbs in this class —not all, though— are mental verbs. (Note that all mental verbs with this argument structure have a non-agentive interpretation.)

[11.c.] illustrates the necessity of the metapredicates CAUSE and CHANGE. According to Komlósy (2000) one default meaning of the ablative case *-tÓl* is the CAUSE of something. Though in cases which are similar to the example above, i.e. where all the three argument structures are well-formed, the change has to be a transition from a state (S) *into* another state (S'). This transition is referred to by the metapredicate CHANGE. There are two arguments to support this thesis. The first relies on the English translation[1]; the elements of this verb class are inclined to be translated into English by perfective verb forms. The structure in sentence [11c] cannot even be put in an imperfective form with the same argument structure. This fact is in accordance with our expectation that sentences with the perfective forms of these structures involve the complete transition between two states, while imperfective forms express the process of transition, but do not imply the end of this process. The other argument takes as its starting point

[1] We have to use English translation, since Hungarian lacks perfective-imperfective distinction expressed by tense.

the observation that there is a verb class with verbs such that the argument with instrumental suffix represents the CAUSE as in the instances above, but there is no transition between definite states which means that CHANGE predicate cannot apply:

[12a] *Az igazgató János - t terhelte a feladat - tal.*
The director John - ACC burden the task - INS
'The director burdened John with the task'

[12b] *A feladat János - t terhelte.*
The task John - ACC burdened
'The task burdened John.'

[12c] *János terhelve van.*
John burdened is
'John is burdened.'

[12d] **János terhelve van a feladat - tól.*
John burdened is the task - ABL
'John is burdened by the task.'

This semantic intuition is caught by the explicit criterion of the syntactical ill-formedness of the sentence [12d]. As the counterexample demonstrates the metapredicate CHANGE is distinctive, that is why we need it independently of CAUSE.

6. Conclusion

Our work aims at creating a well-defined and efficient method for NLP applications to distinguish between verbal complements and adjuncts. The usability of such an algorithm depends basically on two parameters:

— it has to be explicit enough so that different people working on parallel on argument structure descriptions produce coherent, homogeneous work,
— every piece of relevant information that is not predictable by general rules has to be classified as lexical.

We built up our method upon these criteria, assuming that *case suffixes*, the syntactic markers of grammatical functions are not only markers but that there are verb + adjunct structures in which they take part with their own morphosyntactic properties and meanings. These structures are compositional: the verb, the NP and its case suffix form a syntactic unit and the meaning of this unit is calculated adding up the meaning of the verb, the meaning of the NP and that of the case suffix. On the other edge of the scale we find *complements*: they figure in non-compositional structures where the case suffix of the complement NP does not contribute any predictable meaning on its own. The semantic role of the complement NP, i.e. the relation between its denotatum and the action/fact referred to by the verb, depends solely on the verb's lexical properties. Midway between these categories, we found a set of structures in which the case suffix behaves the same way as in adjunction but which are restricted to semantically characterizable classes of predicates. As we would like to reduce the amount of data stored in the lexicon, we decided to capture semi-adjunct structures by the so-called *non-default rules* which only apply to given sets of predicates. These predicates are described by semantic metapredicates.

In accordance with our expectations, the number of 'real' complements reduced considerably. As a secondary result of our work, a cluster of syntactically relevant se-

mantic features is shaping up from the metapredicates that define semantic classes (e.g. cause, change). We expect metapredicates to be language-independent. If this assumption proves to be right, not only default rules but also non-default rules can behave as translation rules for case suffixes: the only modification needed in the rules is the replacement of linguistic labels of NPs in the output of the rule by the target language syntactic marker of the role referred to by the label, e.g.:

rule type	*source language (HUN)*			*target language (EN)*		
non-default	V + change_state	NP	case = ABL	V	NP	prep = 'because of'
default	V	NP	case = INS	V	NP	prep = 'with'

The figure illustrates how a non-default and a default rule may be captured in the bilingual module of a Hungarian-to-English MT system. The non-default rule takes as source language input a verb which denotes a change in someone's state, modified by an NP in ablative case, and translates the case suffix by the preposition *by*. The default rule does not place any restriction on the verb, and states that any NP in instrumental case which has not been matched by earlier rules has to be translated as an instrument, by the preposition *with*.

The most important future task is to find a way to verify the language-independent character of metapredicates. Meanwhile, the precise elaboration of the lexical argument structure database and in parallel, its use in a rule-based machine translation system (Prószéky and Tihanyi 2002) are being carried out. Considering the very strict claim MT sets up for separating language-independent and language-specific information, as well as the practical requirement to cover by rules as many phenomena as possible, we believe that MT as an application is also a relevant test to verify the foundations of our method.

References

Kiefer, F. and Ladányi, M., 2000, «Morfoszintaktikailag semleges képzések». In: *Strukturális magyar nyelvtan* 3. ed. by Kiefer, F., pp. 165-214. Budapest. Akadémiai Kiadó («Morpho-syntactically neutral derivational processes». In: *Structural Grammar of Hungarian* 3.)

Kiss, K., 2002, «The Syntax of Hungarian». Cambridge. Cambridge U. P. Jurafsky, D. and M., J., 2000, *Speech and Language Processing*. Prentice Hall.

Komlósy, A., 1992, «Régensek és vonzatok». In: *Strukturális magyar nyelvtan* 1. ed. by Kiefer, F., pp. 299-529. Budapest. Akadémiai Kiadó («Predicates and Complements». In: *Structural Grammar of Hungarian* 1.)

— 2000, «A műveltetés». In: *Strukturális magyar nyelvtan* 3. ed. by Kiefer, F., pp. 215-291. Budapest. Akadémiai Kiadó («Factitive verbs». In: *Structural Grammar of Hungarian* 3.)

— 2001, *A lexikai-funkcionális grammatika mondattanának alapfogalmai*. Budapest. Tinta kiadó («Syntax in Lexical-Functional Grammar»)

Prószéky, G. and Tihanyi, L., 2002, «Metamorpho: A Pattern-Based Machine Translation Project». In: *Proceedings of the 24th Translating and the Computer Conference*. pp. 19-24. London.

Radford, A., 1988, *Transformational Grammar*. Cambridge. Cambridge U. P.

Silberztein, M., 1993, *Dictionnaires électroniques et analyse automatique de textes: Le systeme Intex*. Paris. Masson.