

# Corpus-driven Study on Hungarian Postpositions

Noémi Ligeti-Nagy

MTA-PPKE Hungarian Language Technology Research Group,  
Práter u. 50/a., 1083 Budapest, Hungary  
`ligeti-nagy.noemi@itk.ppke.hu`

**Abstract.** In this paper, we present a corpus-driven analysis on Hungarian postpositions. Postpositions do not form a clear, well-defined group of words in Hungarian. Although linguistic papers regularly attempted to define strict categories of postpositions, we rather aim to treat these words as a continuum. Six important attributes were chosen based on the literature and corpus queries to define the level of the “postpositionness” of a given word. We created a detailed table of postposition-like elements in Hungarian where every important attribute is represented by a binary value. Based on this table, the morphological annotation of these words can be extended with important features, which can be the basis of their algorithmic syntactic analysis. These features can also be used within the framework of a psycholinguistically motivated, performance-based parser, ANAGRAMMA.

**Keywords:** corpus linguistics, POS-tagging, postpositions

## 1 Introduction

In this paper, we attempt to provide a deep, corpus-driven analysis of the ever-changing group of Hungarian postpositions. Our goal is to look at these words not as a discrete category but rather as a continuum, defined by numerous different features. Our further goal is to analyze them within the framework of a dependency grammar while constantly bearing in mind the specific needs of a psycholinguistically motivated parser (ANAGRAMMA, [1]).

ANAGRAMMA aims to process sentences by following the pattern of human language processing, thus with a strictly left-to-right, word-by-word approach. It works with a *supply* and *demand* framework, meaning that every token has its own *supplies* (or more precisely: a token may have one or more features, and every feature acts as a *supply*) and *demands* (every feature may express one or more *demands*), independently. This requires a rich morphological pre-analysis resulting in a detailed list of features. For example, an article’s *DET* feature expresses a *supply*, which must be used to fulfill the *demand* of a case ending, at the end of a noun phrase. At the end of the utterance, all *demands* should be fulfilled either from the sentence or with default mechanisms. The output of the parser is a dependency graph.

It is rather important to define what we mean by a postposition. From a computational point of view, postpositions are elements unquestionably marking the end of a noun phrase in the sentence (see example 1). However, it requires a deeper literature overview to specify which exact tokens are the members of this group. We are going to somehow characterize these elements based on the literature, although, we aim to provide an algorithm able to analyze these words regardless of whether they are classified as postposition by every linguistic approach or not. We are going to give a list of our criteria in Section 2. In Section 3 and 4, we give an insight to what our corpus study has revealed about postpositions and how they can be described with the help of binary variables indicating the important attributes of this group. In Section 5, we summarize our research.

- (1) *A főnév után jön a névutó.*  
 The noun **after** come the postposition.  
 'The postposition comes after the noun.'

There is hardly another word class that, despite its relatively small number of elements, would generate such a long-lasting debate as Hungarian postpositions. It is almost impossible to find two separate papers arguing the same classification of them. In this section, we attempt to give an overview of what theoretical linguists have said about Hungarian postpositions. In traditional Hungarian grammars [2] postpositions are described as a group of function words forming a morphological-like unit with the noun (phrase) preceding them, therefore functioning as case suffixes. These grammars categorize postpositions as follows:

- (2) a. **Real postpositions**, postpositions taking a caseless noun as a complement: *által* 'by', *alá* 'under.to', *alatt* 'under.at', *alól* 'under.from', *elé* 'in.front.of.to' etc.  
 b. **Postpositions taking a noun with a lexical case**: *fogva* 'as a result of', *kezdve* 'beginning from', *nézve* 'regarding', *keresztül* 'through', *túl* 'beyond', *együtt* 'together' etc.  
 c. **Postpositions with possessive structure**: *alapján* 'based on', *céljából* 'with the aim of', *ellenére* 'despite', *nyomán* 'based on' etc.

This categorization is strongly disputed in papers from a structuralist or generative background. In the third volume of the series *Strukturális magyar nyelvtan* [A Structural Grammar of Hungarian] [3] postpositions are classified into three (and a half) groups:

- (3) a. **Case-like postpositions**: the members of this group act like a case suffix. They do not assign a case to the NP, they take a caseless noun phrase complement. Examples: *által* 'by', *alá* 'under.to' etc.  
 b. **Real postpositions**: elements taking a noun phrase complement with a lexical case: *alul* 'below', *át* 'through', *belül* 'inside of', *kívül* 'outside', *együtt* 'together' etc.

- c. **Other postpositions:** postpositions never co-occurring with a personal pronoun or a demonstrative: *végett* 'with the aim of', *közben* 'during', *óta* 'since', *gyanánt* 'as', *hosszat* 'for'.
- d. (Postposition-like elements): a transitional class containing tokens with a possessive case suffix.

A pure generative approach is still needed. That is provided for example in *The Syntax of Hungarian* by É. Kiss [4]. Here, words earlier described as postpositions are grouped into the following categories:

- (4) a. **Adverbs** taking an argument ("postpositions" taking a noun phrase complement with a lexical case): *együtt* 'together', *alul* 'below' etc.
- b. **Idiomatic participles** taking a noun phrase complement with a lexical case: *nézve* 'concerning', *kezdve* 'beginning from' etc.
- c. **Postpositions** taking a caseless noun phrase complement: *alatt* 'under.at', *mellett* 'near', *révén* 'by means of' etc.

These three divergent categorizations show how uncertain the definition of postpositions in Hungarian linguistic literature is. However, one has to refer to one other paper on this question, namely the doctoral dissertation of Éva Dékány, providing a detailed and sorted list of all Hungarian postpositions [5]. Dékány initiated her categorization from the concept of "naked" and "dressed postpositions". Her inventory is also included in Table 1, where we intended to provide a detailed list of the postpositions ever mentioned in the literature (based on the above discussed four papers).

The main problem of the characterization of postpositions is rather salient: there is a group of postpositions (let us refer to them as "pure postpositions") that always act like postpositions, they bear all the important features of a postposition and finally, they are categorized as postposition by every relevant paper from this field. On the other hand, there are some words, that, and we intended to shed light on this by Table 1, are postpositions from a certain point of view, but are something different from another point of view. Our intention is to capture this continuum: to define the group and the features of prototypical postpositions and circumscribe the layers surrounding this core containing not prototypical postpositions.

Table 1: Encountering all the postpositions mentioned in linguistic papers discussed above. The words in italics are the postpositions. The four columns named MG [2], Str [3], Soh [4] and D.É. [5] indicate the four above-mentioned papers and their opinion on the given postpositions. A number indicates that the given word is categorized as a postposition in the paper. A letter indicates that the given word is mentioned in the paper, but not as a postposition. Different numbers refer to different sub-categorization. Parentheses show that the given word was not mentioned in that paper and we are not certain about its categorization. A “-” means that the given paper does not state anything about that word.

	MG	Str	SoH	D.É.		MG	Str	SoH	D.É.
<i>alá</i> ‘under.to’	1	1	1	1	<i>körül</i> ‘around.at’	1	1	1	1
<i>alapján</i> ‘based on’	3	a	-	(b)	<i>következtében</i> ‘following’	3	a	-	(b)
<i>alatt</i> ‘under.at’	1	1	1	1	<i>közben</i> ‘during’	1	3	-	-
<i>alól</i> ‘under.from’	1	1	1	1	<i>közé</i> ‘between.to’	1	1	1	1
<i>által</i> ‘by’	1	1	1	1	<i>közel</i> ‘close to’	2	-	(a)	2
<i>alul</i> ‘below’	2	2	a	2	<i>között</i> ‘between.at’	1	1	1	1
<i>át</i> ‘through’	2	2	a	2	<i>közül</i> ‘between.from’	1	1	1	1
<i>belül</i> ‘inside of’	2	2	a	2	<i>léte</i> ‘despite being’	1	a	-	b
<i>céljából</i> ‘with the aim of’	3	a	-	(b)	<i>mellé</i> ‘near.to’	1	1	1	1
<i>együtt</i> ‘together’	2	2	a	2	<i>mellett</i> ‘near.at’	1	1	1	1
<i>elé</i> ‘in.front.of.to’	1	1	1	1	<i>mellől</i> ‘near.from’	1	1	1	1
<i>ellen</i> ‘against’	1	1	1	1	<i>miatt</i> ‘because of’	1	1	1	1
<i>ellenére</i> ‘despite’	3	a	-	1	<i>módjára</i> ‘way.of’	1	a	-	-
<i>elől</i> ‘in.front.of.from’	1	1	1	1	<i>módra</i> ‘mode.of’	1	-	-	-
<i>előtt</i> ‘in.front.of.at’	1	1	1	1	<i>mögé</i> ‘behind.to’	1	1	1	1
<i>esetén</i> ‘in case of’	3	a	-	1	<i>mögött</i> ‘behind.at’	1	1	1	1
<i>felé</i> ‘towards’	1	1	1	1	<i>mögül</i> ‘behind.from’	1	1	1	1
<i>felől</i> ‘from the direction of’	1	1	1	1	<i>multán</i> ‘after (time)’	1	-	-	-
<i>felül</i> ‘over’	2	-	(a)	2	<i>múlva</i> ‘after (time)’	1	-	-	1
<i>fogva</i> ‘as a result of’	2	2	b	a	<i>nélkül</i> ‘without’	1	1	1	1
<i>fölé</i> ‘above.to’	1	1	1	1	<i>nézve</i> ‘regarding’	2	2	b	a
<i>fölött</i> ‘above.at’	1	1	1	1	<i>nyomán</i> ‘based on’	3	a	-	1
<i>fölül</i> ‘above.from’	2	-	(a)	1	<i>óta</i> ‘since’	1	3	(1)	1
<i>folytán</i> ‘as a consequence of’	3	-	-	1	<i>része</i> ‘for’	3	a	-	b
<i>gyanánt</i> ‘as’	1	3	(1)	1	<i>révén</i> ‘by means of’	3	a	-	b
<i>hasonlóan</i> ‘similarly to’	-	-	-	2	<i>során</i> ‘in the course of’	3	a	-	1
<i>helyett</i> ‘instead of’	1	1	1	1	<i>számára</i> ‘for’	3	a	-	b

<i>hosszat</i> 'for'	-	3	-	-	<i>szembe</i> 'opposite.to'	(2)	-	(a)	2
<i>innen</i> 'on this side of'	2	-	-	2	<i>szemben</i> 'opposite.at'	2	-	(a)	2
<i>iránt</i> 'towards'	1	1	1	1	<i>szemből</i>	(2)	-	(a)	2
					'opposite from'				
<i>ízben</i> 'times'	1	-	-	-	<i>szemközt</i>	(2)	-	(a)	2
					'opposite at'				
<i>javára</i> 'in favour of'	3	a	-	b	<i>szerint</i> 'according to'	1	1	1	1
<i>kedvéért</i>	3	a	-	b	<i>tájban/tájt</i> 'around'	1	-	-	1
'for the sake of'									
<i>képest</i> 'compared to'	-	-	(a)	2	<i>túl</i> 'beyond'	2	2	a	2
<i>keresztül</i> 'through'	2	2	a	2	<i>túlra</i> 'beyond.to'	(2)	-	(a)	2
<i>kezdve</i>	2	2	b	a	<i>túlról</i> 'beyond.from'	(2)	-	(a)	2
'beginning from'									
<i>kivéve</i> 'except for'	(2)	-	(b)	a	<i>után</i> 'after'	1	1	1	1
<i>kívül</i> 'outside'	2	2	a	2	<i>útján</i> 'by way of'	3	a	-	1
<i>kívülre</i> 'outside.to'	(2)	-	-	2	<i>végett</i>	1	3	(3)	1
					'with the aim of'				
<i>kívülről</i> 'outside.from'	(2)	-	-	2	<i>végig</i> 'to the end of'	2	-	(a)	2
<i>köré</i> 'around.to'	1	1	1	1					

## 2 Methods, materials

In Section 1, we stated that postpositions' importance - from a computational point of view - lies in them indicating the strict end of a noun phrase just as case suffixes do. However, numerous members of Table 1 tend to appear in other positions both before and after the noun phrase. We, in this study, as our further goal is an algorithmic description of the behavior of postposition-like elements, proceed by keeping only one of the above-mentioned criteria: the candidate word has to take a noun phrase, caseless or not, as a complement.

One section of the research triggered the next; we initiated our searches from the list of the words in Table 1, but continuously expanded it with other postposition candidates popping up in the queries.

The corpus used is the 2.0.4. version of the *Hungarian Gigaword Corpus* [6].

Based on the literature, and on the results of our numerous queries, the following important features were outlined:

- **position**: by position we mean the preferred succession of the postposition and its complement regardless of their adjacency.
- the **case-marking** of the complement: at first, to keep the features binary, we started by differentiating postpositions with a caseless noun and postpositions taking a noun with a lexical case. However, later on, it will be necessary to distinguish the postpositions taking a noun with a lexical case by their required case.
- **adjacency**: Are the postposition and the noun always strictly adjacent, or other tokens may intersect between them?

- their position in **wh-questions**: do the postposition always follow the wh-word (see example 5a), or can it stay behind (5b)?
- (5) a. ***Ki** után jövök?*  
 Who after come.Sg1?  
 'After whom do I come?'
- b. ***Min** mentél keresztül?*  
 What.SUP go.PastSg2 through?  
 Through what did you go?'

These features are mainly syntactical, and especially motivated by the computational point of view that we apply here. However, we have to encounter some morphological typicality that does not influence the computational analysis of these words but is important in order to have a complex view of them.

- **demonstratives**: is the postposition copied onto the demonstrative as well (see example 6a) or only the case marker is (6b)?
  - person-number **agreement**: when postpositions take a pronominal complement, where does the agreement marker appear? On the postposition itself (example 7a), or on another element (7b)?
- (6) a. *az alól a rét alól*  
 that below.from the meadow below.from  
 'from under that meadow'
- b. *azon a réten át*  
 that.SUP the meadow.SUP through  
 'through that meadow'
- (7) a. *alólam*  
 under.from.Sg1  
 'from under me'
- b. *rajtam keresztül*  
 on.Sg1 through  
 'through me'

### 3 Results

Our results are summarized in Table 2. Table 2 shows the analysis of postpositions based on features of binary values. Column “pos” evaluates the position of the postposition relative to the noun: if the postposition always follows the noun, the value here is 1; otherwise 0. Column “NOM” is related to the case invoked by the postposition: if the noun is always caseless, the value here is 1, otherwise

0. In column “adj”, we evaluate the adjacency of the two words: if the noun and the postposition are always next to each other, the value is 1, otherwise 0. The column “wh” refers to the behavior of the postposition in wh-questions: if the postposition follows the wh-word as its postposition, then the value is 1 (see example 5a), otherwise 0 (example 5b). Column “dem” represents postpositions’ behavior with demonstrative pronouns: if the postposition is copied onto the demonstrative as well, the value is 1 (see example 6a), otherwise 0 (example 6b). Finally, column “pers pron” contains the value 1, if the person-number agreement appears on the postposition itself (example 7a), 0 otherwise (example 7b).

It must be noted regarding the methodology, that five counterexamples were needed to prevent a given postposition from receiving a value of 1 for a specific property. For example, if the word appears before a noun five times, then its value for the feature “pos” is 0. Therefore, corpus queries used to build the database presented in Table 2 were mainly searches to prove the existence of counterexamples: if the query resulted in four or fewer matches, the given postposition received a value of 1 for the given feature.

The evaluation of the “wh” property based on the corpus is particularly difficult, therefore in most cases, the value was determined based on our linguistic intuition. Cells that contain a ?, indicate that the acceptability of our examples of testing a given property is not entirely certain.

Table 2: Encountering all the postpositions from the literature and their attribute values. A value of 1 indicates that the given postposition always produces the prototypical behavior of postpositions in the syntactic structure under examination. Column *pos* is about the position of the word with regard to the noun. Column *NOM* represents whether the postposition always takes a caseless noun. Column *adj* is about the strict adjacency of the two words. Column *wh* represents the word’s behavior in wh-questions. The two columns after the vertical line are the morphological attributes: *dem* is about the structure with a demonstrative pronoun, *pers pron* about the structure with a personal pronoun.

postposition	meaning	pos	NOM	adj	wh	dem	pers pron
<i>alá</i>	under.to	0	0	1	1	1	1
<i>alapján</i>	based on	1	0	1	1	0	0
<i>alatt</i>	under.at	1	1	1	1	1	1
<i>alól</i>	under.from	1	1	1	1	1	1
<i>által</i>	by	1	0	1	1	1	1
<i>alul</i>	below	1	0	0	1	0	0
<i>át</i>	through	0	0	0	0	0	0
<i>belül</i>	inside of	?	0	1	1	0	0
<i>céljából</i>	with the aim of	1	0	1	1	0	-
<i>együtt</i>	together	0	0	0	0	0	0

<i>elé</i>	in.front.of.to	0	0	1	1	1	1
<i>ellen</i>	against	1	1	1	1	1	1
<i>ellenére</i>	despite	1	0	1	1	0	-
<i>elől</i>	in.front.of.from	1	1	1	1	1	1
<i>előtt</i>	in.front.of.at	1	1	1	1	1	1
<i>esetén</i>	in case of	1	0	1	1	0	-
<i>felé</i>	towards	1	1	1	1	1	1
<i>felől</i>	from the firection of	1	1	1	1	1	1
<i>felül</i>	over	1	0	0	0	0	0
<i>fogva</i>	from (time)	1	0	1	1	0	0
<i>fölé</i>	above.to	1	1	1	1	1	1
<i>fölött</i>	above.at	1	1	1	1	1	1
<i>fölül</i>	above.from	1	0	1	1	0	1
<i>folytán</i>	as a consequence of	1	0	1	?	0	-
<i>gyanánt</i>	as	1	1	1	-	-	-
<i>hasonlóan</i>	similarly	0	0	0	1	0	0
<i>helyett</i>	instead of	1	1	1	1	1	1
<i>hosszat</i>	for	1	1	1	-	-	-
<i>innen</i>	on this side of	0	0	0	?	0	0
<i>iránt</i>	towards	1	1	1	1	1	1
<i>ízben</i>	times	1	1	1	-	-	-
<i>javára</i>	in favour of	1	0	0	1	0	1
<i>kedvéért</i>	for the sake of	1	0	0	1	0	1
<i>képest</i>	compared to	1	0	0	1	0	0
<i>keresztül</i>	through	0	0	0	0	0	0
<i>kezdve</i>	beginning from	0	0	1	1	0	0
<i>kivéve</i>	except for	0	0	0	1	0	0
<i>kívül</i>	outside	0	0	0	1	0	0
<i>kívülre</i>	outside.to	1	0	0	1	0	0
<i>kívülről</i>	outside.from	1	0	1	1	0	-
<i>köré</i>	around.to	1	1	1	1	1	1
<i>körül</i>	around.at	1	1	1	1	1	?
<i>következtében</i>	as a consequence of	1	0	1	1	0	-
<i>közben</i>	during	1	1	1	1	1	-
<i>közé</i>	between.to	1	1	1	1	1	1
<i>közel</i>	close to	0	0	0	0	0	0
<i>között</i>	between.at	1	1	1	1	1	1
<i>közül</i>	between.from	1	1	1	1	1	1
<i>létére</i>	despite being	1	1	1	-	-	-
<i>mellé</i>	near.to	1	1	1	1	1	1
<i>mellett</i>	near.at	1	1	1	1	1	1
<i>mellől</i>	near.from	1	1	1	1	1	1
<i>miatt</i>	because of	1	1	1	1	1	1
<i>módjára</i>	way of	1	1	1	-	0?	-
<i>módra</i>	mode of	1	1	1	-	-	-



<i>mögé</i>	behind.to	1	1	1	1	1	1
<i>mögött</i>	behind.at	1	1	1	1	1	1
<i>mögül</i>	behind.from	1	1	1	1	1	1
<i>múltán</i>	after (time)	1	1	1	1	0	-
<i>múlva</i>	after (time)	1	1	1	-	-	-
<i>nélkül</i>	without	1	1	1	1	1	1
<i>nézve</i>	regarding	1	0	1	1	0	0
<i>nyomán</i>	based on	1	0	1	1	0	-
<i>óta</i>	since	1	1	1	1	1	-
<i>részére</i>	for	1	0	1	1	0	1
<i>révén</i>	by means of	1	0	1	1	0	-
<i>során</i>	in the course of	1	0	1	1	0	1
<i>számára</i>	for	1	0	1	1	0	1
<i>szembe</i>	opposite.to	0	0	0	1	0	0
<i>szemben</i>	opposite.at	0	0	0	1	0	0
<i>szemből</i>	opposite.from					0	0
<i>szemközt</i>	opposite.at	0	0	0	1	0	0
<i>szerint</i>	according to	1	1	1	1	1	1
<i>tájban/tájt</i>	around (time)	1	1	1	-?	-?	-
<i>túl</i>	beyond	0	0	0	0	0	0
<i>túlra</i>	beyond.to	1	0	0	1	0	0
<i>túlról</i>	beyond.from	1	0	0	1	0	0
<i>után</i>	after	1	1	1	1	1	1
<i>útján</i>	by way of	1	0	1	1	0	-
<i>végett</i>	with the aim of	1	1	1	1	-	1
<i>végig</i>	'to the end of'	0	0	0	0	0	0

## 4 Discussion

The first thing to see in Table 2 is that “-” is very frequent in columns 4-6, indicating that the given postposition does not appear in the structure under examination. For example, the word *gyanánt* ‘as’ cannot be connected to a personal pronoun, therefore in the last position of the vector of *gyanánt* a “-” can be seen.

If we concentrate on the postpositions with a vector containing only 1 values, the group of prototypical postpositions is outlined: these are almost completely identical to the group of “pure postpositions”; the words categorized as postpositions by every linguistic paper. The exceptions are, on the one hand, the postpositions the base form of which is homonymous to the one attached to a third person singular personal pronoun (see examples 8a and 8b). These tokens may appear in front of the noun as well (example 9). Another exception is *által* ‘by’, which sometimes - mainly in the literary subcorpus, but also in a small number in the personal subcorpus - takes a noun with a lexical case (example 10).

- (8) a. *elé*  
in.front.of  
'in front of'
- b. *elé*  
in.front.of.Sg3  
'in front of him/her'
- (9) ... *szólt elé*                      *a kocsisnak.*  
... said in.front.of the driver.DAT.  
'... said to the driver.'
- (10) *Természetesen szintén szigorúan tévén*                      *által.*  
Naturally as.well strictly television.SUP by.  
'Naturally, also strictly by television.'

If we omit the 4-6. values of the properties, another significant group emerges: the words which received a value of 1 for the first three, syntactic properties. Since the last three features as conditions cannot be fully applied to these words, omitting those should not be a problem in an algorithmic processing of these tokens. However, the value 1 in the first three cells of the vector of these words indicates that they would be worth annotated in the corpus as postpositions, since they always take the final position in a noun phrase, strictly following a noun without a lexical case.

The odd one out in the table is *szemből* 'opposite from': this word should not be considered a postposition in any way since it can not even get values for the key syntactic properties as it does not occur in syntactic structures like that.

The group of words on the periphery is also clearly outlined: they are the ones receiving only 0-s. Their common feature is that they can be examined from every aspect, meaning that they do occur in every syntactic structure in which typical postpositions do, however, they behave differently than those. In their annotation it should be worth sticking to their adverbial character; they should be annotated as adverbs taking an argument, which may precede or follow them and may appear further away in the text.

The most interesting elements in Table 2 are the ones represented by a vector beginning with 1 0 1 (disregarding the values in the 4-6. cells): these are postpositions always strictly following a noun with a lexical case. Still, they are closer to prototypical postpositions than adverbs are. Their analysis must differ from this latter group (because of the noun bearing a case suffix in contrast with prototypical postpositions, where the noun is caseless), and from that of the adverbs as well (because they are not looking for a complement anywhere in the sentence, theirs always precedes them). In this paper, this group will be referred to as postpositions taking an argument.

Table 3 shows tokens often appearing in our corpus queries with postposition-like behavior, some of which has already been mentioned as postposition-candidates

**Table 3.** List of postposition-like elements and their feature vectors popped up in our corpus queries.

postposition	meaning	pos	NOM	adj	wh	dem	pers	pron
<i>címén</i>	in the name of	1	0	1	1	0	-	
<i>dacára</i>	despite	0	0	1	1	0	-	
<i>eltérően</i>	differently	0	0	1	1	0	0	
<i>esetében</i>	in case of	1	0	1	1	0	1	
<i>érdekében</i>	in favor of	1	0	1	1	0	1	
<i>értelmében</i>	according to	1	0	1	1	0	-	
<i>fényében</i>	in view of	1	0	1	1	0	-	
<i>függően</i>	depending on	1	0	1	1	0	0	
<i>hiányában</i>	in default of	1	0	1	1	0	-	
<i>idején</i>	in	1	0	1	1	0	-	
<i>jegyében</i>	in spirit of	1	0	1	1	0	-	
<i>keretében</i>	within the framework of	1	0	1	1	0	-	
<i>kezdődően</i>	beginning from	1	0	1	1	0	0	
<i>köszönhetően</i>	thanks to	0	0	1	1	0	0	
<i>követően</i>	following	1	0	1	1	0	0	
<i>megegyezően</i>	same way as	1	0	1	1	0	0	
<i>megelőzően</i>	previous to	1	0	1	1	0	0	
<i>megfelelően</i>	accordingly	0	0	1	1	0	0	
<i>terén</i>	in the field of	1	0	1	1	0	-	
<i>ürügyén</i>	under cover of	1	0	1	1	0	-	
<i>vonatkozásán</i>	with respect to	1	0	1	1	0	0	

in other papers [7]. As can be seen, strict adjacency is a common feature of them; furthermore, almost every one of them appears exclusively in a noun phrase ending position, after the noun. Therefore, they are close relatives of the group of postpositions taking an argument. The difference between the two groups is that these postposition-candidates have a more complex morphological structure; they contain a possessive case marker or an essive case suffix. Their syntactic analysis is not much different from what their detailed morphological analysis would activate, but their meaning and their exclusive occurrence in this typical position of postpositional elements justify them to be included in the group of postpositions.

Based on the aforementioned results, three major groups of postposition-like elements are outlined.

- prototypical postpositions: these words can always be found directly after a noun without a lexical case. Our suggestion is to use **POSTP** as their POS-tag – as it is already the case with most of the words of this group. The members of this group are the words with a vector beginning with 1 1 1 in Table 2.
- for words in the periphery (compared to the prototypical ones): these words always take a noun with a lexical case as an argument; they can appear in front of the noun as well, and other elements can intersect between them. They received a value of 0 in every cell in Table 2. Our suggestion is to use

ADV as their POS-tag, and provide them with a feature to indicate which lexical case they take.

- for postpositions taking a noun with a lexical case: although occurring always strictly after the noun, these words differ from prototypical postpositions as they take a noun with a lexical case. In Table 2 and Table 3 their vectors start with 1 0 1. Their algorithmic analysis requires further study.

## 5 Conclusion

In this paper, a detailed, corpus-driven analysis of the Hungarian postposition-like elements was presented. We examined how six features mentioned in linguistic literature, used for the distinction of the disjunct categories of postpositions, characterize these words when studied in a corpus. As can be seen in Table 2, the numerous postposition-candidates can be clearly arranged based on these features. We found prototypical and peripheral elements, the algorithmic analysis of which must be different. Three main groups were outlined with three algorithms proposed for their syntactic analysis. In the next step of our research, the implementation of these algorithms must be followed by an evaluation. It is also important to assign frequency numbers to the 0 values in Table 2, to measure significance and to further refine the classification of postpositions.

## References

1. Prószéky, G., Indig, B.: Magyar szövegek pszicholingvisztikai indíttatású elemzése számítógéppel [Psycholinguistically motivated analysis of Hungarian texts with computer]. *Alkalmazott Nyelvtudomány*. 15, 29–44 (2015)
2. Keszler, B.: Magyar grammatika [Hungarian grammar]. Nemzeti Tankönyvkiadó, Budapest (2000)
3. Kiefer, F., Bánréti, Z.: Stukturális magyar nyelvtan: Mondattan. [A Structural Grammar of Hungarian: Syntax]. Akadémiai Kiadó, Budapest (1992)
4. É. Kiss, K.: The Syntax of Hungarian. Cambridge Syntax Guides. Cambridge University Press (2002)
5. Dékány, É. K.: A profile of the Hungarian DP: the interaction of lexicalization, agreement and linearization with the functional sequence. PhD dissertation. University of Tromsø (2012)
6. Oravecz, Cs., Váradi, T., Sass, B.: The Hungarian Gigaword Corpus. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC '14)*, pp. 1719–1723. European Language Resources Association (2014)
7. Ligeti-Nagy, N.: Szövegtörzsek pontosabb annotációja gépi elemzéshez [Improving corpus annotation for computational parsing]. In: *Többsz nyelvűség és kommunikáció Kelet-Közép-Európában. XXIV. Magyar Alkalmazott Nyelvészeti Kongresszus*, pp. 421–429. Erdélyi Múzeum-Egyesület (2015)