

1 Introduction

The following text is intended as supplementary material for the article *Towards capturing implicit innovative language attitude using an auditory Implicit Association Test*. It contains a more detailed structural description of the variables tested in Experiment 1,¹ including the results of frequency studies on the basis of corpus queries, which served as a background for the interpretation of the test results. Besides, the labels “more frequent” and “less frequent”, which were assigned to the variants to be able to distinguish them systematically, are also based on the results of the corpus queries. However, it is also necessary to add that we did not mean to overestimate the significance of the search results, that is, we do not claim that this opposition between the forms faithfully represents “real” language use. Still, these simple labels will hopefully be useful for practical purposes (i.e., consistent differentiation of the two variants of a variable).

The test used for eliciting grammaticality judgments (and the possible evaluation differences between the different age groups) contained 11 structures that exhibit variation in Modern Standard Hungarian (abbreviated as MSH in the following). As many details of the history of Hungarian are thoroughly researched, most of these variants could be classified reliably as “innovative” or “conservative” on a diachronic basis. However, some of these variants have co-existed for hundreds of years, which would make it rather questionable to label them mechanically “innovations” or “archaisms” on the basis of their diachronic background within the framework of the present study investigating synchronic language use. In what follows, we will discuss each of the tested linguistic variables in more detail, with respect to the following guidelines: the structural description of the given form, its diachronic background, and data concerning the use of the variants, obtained on the basis of queries of the Hungarian National Corpus (HNC; for a brief description, see below). Hopefully, these different types of background information will help the interpretation of test data.

The new version of the HNC (Oravecz et al. 2014) is a large, modern corpus of contemporary Hungarian, equipped with thorough linguistic annotation and a fast and detailed corpus query interface. The annotation contains morphological information (inflection, derivation, compounding, morphemes) and phonological information (natural classes, CV skeleton) as well. The currently available part of it contains six genres in the following proportions: journalism 35.5%, literature 6.6%, (popular) science 20.1%, personal (web forums) 7.9%, official 16.9%, transcribed spoken (radio) 13.9%.² With respect to *modality* (speech vs. written text), the *speaking style* (or in other words, spontaneity, that is, the timing of speech planning processes, and the articulation of the utterances) and *register* (specifically the formal-informal continuum), the following features are considered to be characteristic of these subcorpora.

“Journalism” contains only written material that can be considered to be closer to the “formal” end of the continuum. “Literature” is also compiled only of written material and contains texts that, naturally, intertwine formal and informal registers. Similarly to “literature”, “(popular) science” contains written texts, but more to the formal end of the

¹ For the description of Experiment 1, see Section 3 in the article.

² To the best of our knowledge, this subcorpus of HNC is the largest morphologically annotated corpus of spoken Hungarian.

formal-informal continuum. “Personal” is compiled of written texts (posts of web forums) that may be positioned, again, more to the informal end of the continuum. “Official” refers to transcribed speeches that were written in advance and read aloud at the Parliament (this is the bigger part of the subcorpus), but this also contains some spontaneous speech that was improvised on the spot. “Transcribed spoken (radio)” is a subcorpus in which mainly transcribed spontaneous speech and a smaller amount of read aloud speech can be found that was originally recorded on a country-wide radio station (Kossuth Rádió). The registers used in these texts may be considered mixed, again, but more to the end of the formal continuum: although informal utterances are also observable in this subcorpus, it is clearly noticeable that both the (editorial) staff and the guests are aiming at producing more elevated types of speech. It is important to note that simultaneously informal and spontaneous speech is essentially missing from the HNC, and this might be the cause of the underrepresentation of the supposedly innovative variants analyzed and investigated in the study. In what follows, we summarize the results of the corpus analysis, and we refer to the tested phenomena (listed in Section 3.1 of the article) in the heading of each subsection.

2 Article drop (phenomena 1, 2, 3)

The Hungarian definite article *a/az* emerged during the Old Hungarian period through the functional split of the distal demonstrative pronoun *az* ‘that’, a grammaticalization change that was attested in various languages (Heine & Kuteva 2004, Harris & Campbell 1995, Givón 2001). Its use displays gradual extension during Old Hungarian, appearing in more and more grammatical contexts where it marks definiteness (I. Gallasy 1991, 1992, Egedi 2014).

However, in the test we looked at examples which could (theoretically) be interpreted as a change in the opposite direction:³ the article does not appear in environments in which it would be compulsory, or at least preferable in MSH. An example of this type is shown in (1), observed by Nádasy (2008).⁴

- (1) *Nyugatiba* *bejössz* [...]

Western-ILL drop.in.PRS.2SG

‘You come into the Western (Railway Station)’

With respect example (1), Nádasy claims that while this form does not conform to MSH (and one reason for this is the lack of the article before *Nyugati*), it is still generally observable in spoken language, and the environment that might trigger article drop is the sentence-initial position of the noun. In addition to the position of the noun, we decided to test for an additional variable, namely whether names of institutions like “Nyugati” in (1) show a tendency to be used without a definite article.

³ However, it is almost impossible to investigate whether in these specific instances the use of the article had already been general, and the drop of the article is a further change, or (more likely) there has always been a significant amount of variation, and, due to so far unknown reasons, the more archaic, article-less pattern started to spread in MSH.

⁴ ACC = accusative, DAT = dative, DEL = delative, ILL = illative, INF = infinitive, INS = instrumental, MOD = marker of epistemic modality; PL = plural, PRS = present, PST = past, PV = preverb (marking telicity), SUP = superessive.

Concerning sentence-initial position, the query of HNC ran with the following settings. First we searched for plural nouns in the nominative in sentence initial position, with a further restriction that the noun cannot be suffixed with possessive markers (example (2) below is a typical example of this pattern).⁵

- (2) (az/0) *ajtók a bal oldalon nyílnak.*
 (the/0) door.PL the left side.SUP open.PRS.3PL
 ‘Doors open on the left.’

The number of hits when these nouns occurred with and without the definite article was 7,480 and 14,430, respectively. For a more detailed analysis, we have chosen a random list of one thousand hits of the latter group. It turned out that the overwhelming majority of the article-less items were titles (appearing in newspapers, scientific works etc.), which do show a tendency to drop articles owing to the specific style they represent. The rest of the examples cannot get a uniform explanation: even if we could assume (on the basis of our Hungarian competence) that certain structures can have competing variants that are identical except for the presence of a determiner, the competition is not always between a definite determiner and a null, but can just as well be a competition between an indefinite determiner and a null. Therefore, as there are very few genuine examples of article drop,⁶ we could not attest the general nature of sentence-initial article deletion on the basis of HNC, which, naturally, can be due to several reasons, including the fact that spontaneous informal language use is underrepresented in the corpus, as noted above.

As for names of institutions, in MSH these occur with the definite article (as an indirect evidence, the comprehensive handbook of normative linguistics explicitly states this, cf. Grétsy & Kovalovszky 1980, 104). However, on the basis of our own observations we had the impression that some speakers prefer to drop the article before these. To obtain a larger number of data (and, as in this case the hits did not need much manual analysis), we searched the entire corpus for all the instances for the verbs *bemegy* ‘enter’ and *bengrik* ‘stop by’ with an illative-marked noun starting with a capital letter in its immediate environment. The query resulted in 220 hits in which the noun had an article and no hits for article-less nouns. Therefore, the conclusion is similar to that of the previous case: article drop before names of institutions is not yet attestable in HNC, either because of the novelty of this phenomenon, or due to nature of the corpus.

A final test point was to see article use with nouns in sentence medial position. Naturally, this is too broad a category to be tested in general, therefore, we selected sentences to be tested in which we happened to observe article drop ourselves (e.g., the article-less version of example (3) was attested by us in spontaneous communication). In the test these sentences occurred with and without the article (see 3.1 in the article), and the focus of the corpus queries was finding the distribution of these two patterns.

⁵ This restriction was necessary as possessive-marked nouns, which are predominantly interpreted as definite in themselves, differ significantly in this respect. In order to illustrate this, we looked for possessive marked nouns in sentence initial position as well. In this case we did not perform a manual analysis of the findings, but it is telling that there were 6619 hits with an article, and about 87,000 hits without an article.

⁶ E.g., “*Kezddők a Standard Scant bajtsák végre, a haladóknak viszont lehetőségük van kizáró listát készíteni.*” ‘Beginners should execute standard scan, however, the advanced ones have the option to make a disqualifying list.’

- (3) *Sajnos* (a) *kóruossal* *lesz* *fontos* *prób-ám*.
 Unfortunately (the) choir.INS will.be.3SG important rehearsal.1SG
 ‘Unfortunately, I will have an important rehearsal with (the) choir.’

The test itself aimed at investigating the use (and potential drop) of the definite article before nouns with and without possessive suffixes, therefore, we also tried to find parallels for the investigated structures from HNC. First, we looked for sentences similar to (3) in the corpus: those starting with the sentence adverbial *sajnos* ‘unfortunately’, followed by nouns with an instrumental marker, the latter either preceded by an article, or immediately following the sentence adverbial. There were 97 hits with the article and 67 without it (i.e., adjacency of the adverbial and the noun), but the manual analysis of these latter cases showed that none of these could be taken as examples of article drop (i.e., the noun in the instrumental could not have a determiner in the given context, or, if it could have some determiner, it would not necessarily be the definite article). Therefore, our observation cannot be confirmed by corpus data. Nevertheless, these findings justify referring to the article-less variant as “rare”, similarly to the other hypothesized cases of article deletion.

In contrast, we looked at instances of nouns meaning ‘sibling’ (*öcs* ‘younger brother’, *báty* ‘older brother’, *húg* ‘younger sister’, *nővér* ‘older sister’) marked with a first person singular possessive suffix (*-m*) in HNC. The specific patterns we looked for was verb + definite article + possessive marked noun vs. verb + possessive-marked noun appearing adjacently (again, we had similar structures in the test). In this case, there were 481 hits for the pattern with the article, and 219 for the pattern without the article. The manual analysis of the latter group showed that there were 112 structures that can be considered as cases of article drop, i.e., when the noun could have a definite article (and, more than likely, it could only have a definite article), but the speaker chose not to use it. It seems to be the case that possessive-marked kinship terms can freely occur without the definite article, still, this pattern occurs less often than the other pattern.

3 Presence or absence of the subordinator *hogy* (phenomena 4, 5, 6, 7, 8)

The most general subordinator in Hungarian is *hogy* ‘that’, a complementizer that developed presumably during the Proto-Hungarian period through the reanalysis of the question word *hogy* ‘how’ in embedded questions or, alternatively, of the pronominal adverbial *hogy* ‘as; the way that’, which were homophonous at the time of this change (in MSH, the latter is *ahogy*; for a general description of the grammaticalization process, see Haader 1991). The investigations of the earliest sources reveal that by Old Hungarian the complementizer appears in all of those functions that it has in MSH (Haader 1995, Bácskai-Atkári & Dékány 2014). However, it was not an obligatory marker of finite subordination then, and its use is not obligatory in MSH, either. Summarizing the typical instances when the use of *hogy* is optional, Kenesei (1992, 673-679) distinguishes external and internal conditions of *hogy*-deletion. The former category covers conditions pertaining to the grammatical relationship of the main sentence and the subordinate sentence, while the latter category encompasses features that characterize the given subordinate sentences headed by the complementizer *hogy*.

Kenesei (1992, 674) points out that the external criteria of complementizer deletion are reducible to a single principle: deletion is only possible if the subordinate clause is properly governed by the verb of the matrix clause, i.e., the verb of the matrix clause

assigns theta-role to the subordinate clause, and the verb and the subordinate clause have to occupy sister nodes. For instance, this excludes cases when the matrix clause contains structural focus or negation, as in these cases the verb obligatory moves to a higher structural position, therefore it cannot be adjacent to the complementizer. The complementizer cannot be omitted in those cases when the subordinate clause precedes the main clause, either, and the reason behind this restriction is the same: the verb of the main clause and the complementizer of the subordinate clause cannot be adjacent in that pattern.

Concerning inner criteria, one of them (that overrides all principles listed above) is that the complementizer *hogy* can be omitted freely if the subordinate clause is an embedded question. Another criterion is the modality of the subordinate clause: if it is an embedded imperative, the complementizer can be deleted as well. Furthermore, if there is a second subordinate clause embedded into the subordinate clause headed by *hogy*, and, as a result of this, the two complementizers are adjacent, some complementizers facilitate the deletion of *hogy*. Finally, it is necessary to mention that there must be individual differences between the speakers concerning *hogy*-deletion, and stylistic rules can also influence this phenomenon. Most Hungarians learn at school that it is “not elegant” to iterate the complementizer *hogy* when embedded clauses follow each other, and in certain cases the written samples show omission even in cases when its grammaticality is dubious.

Out of the three sections that cover the presence or absence of *hogy* in the test (phenomena 4, 5-7, 8 in Section 3.1), one set contains examples where the complementizer could be freely omitted in MSH, as in example (4) below.

- (4) *Azt hittük, (hogy) itt tilos a dohányzás.*
 that.ACC think.PST.3SG<1PL (that) here forbidden the smoking.
 ‘We thought (that) smoking was forbidden here.’

In order to acquire data that could facilitate the interpretation of the test results (presented in Section 3.3 of the article), there were three verbs (potentially subcategorizing for *hogy*-clauses) that were taken as the starting points of corpus query in HNC. We searched for the occurrences of the verbs *mond* ‘say’,⁷ *hisz* ‘think, believe’, *bevall* ‘admit’ in the objective conjugation, as in MSH (as opposed to earlier periods) those verbs that properly govern the subordinate clause cannot be in the subjective conjugation. The search results were filtered for subcorpora, and a random sample of 250 items of the six subcorpora was generated for each of these three verbs, so altogether we looked at 1,500 occurrences of each verb. These were then analyzed manually for the feature investigated (overt complementizer, deleted complementizer, and irrelevant cases). Figure 1 shows the results of this analysis.⁸

The data reveal that there does not seem to be a clear pattern concerning the deletion of *hogy*: neither the verbs nor any of the subcorpora show characteristic distribution with respect to this option. We do not mean to draw the conclusion that there would not be any tendencies governing the choice of the speakers to use or drop

⁷ In the case of *mond*, we excluded those occurrences that were in the first person plural (*mondjuk* ‘we say’ or ‘let us say’), as this functions as a discourse marker quite frequently.

⁸ These cover those items in which either the given verb had no subordinate clause of the relevant type, or in which there was some feature that generally blocks *hogy*-deletion, e.g., when the subordinate clause precedes the main clause.

the subordinator, only that these tendencies could not be captured through this search. The sample, as usually, could have been larger, and the query could have been conducted with more specified grammatical features (e.g., generating samples in which either all or none of the matrix sentences contain an overt pronoun associated with the subordinate clause), but it is quite likely that the differences between the speakers and random stylistic choices would raise difficulties even in that case. Therefore, when labeling the variants in phenomenon 8 (optional *hogy*-deletion), using “more frequent” to mark the variant with the subordinator is an arbitrary choice.

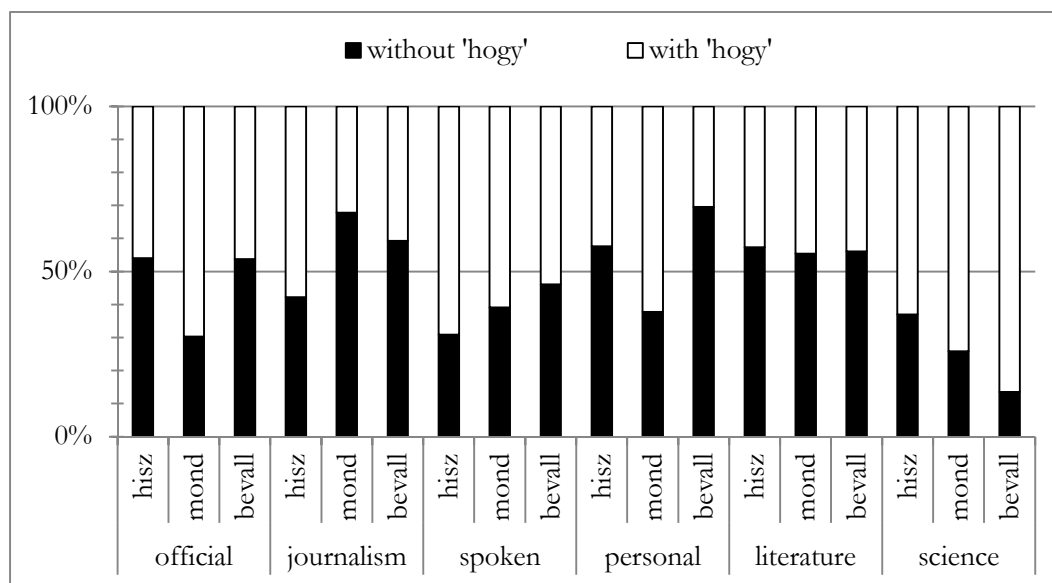


Figure 1: The proportion of subordinate clauses with and without the subordinator *hogy*

In the section of the test that focused on the use of *hogy*, another set contained sentences in which the verb of the main clause was *lehet* (‘it may be’). In contrast with the verbs investigated above, the subordinate clause of this matrix verb would necessarily be headed by an overt complementizer (i.e., *hogy*) in MSH. Still, especially in non-formal registers, *hogy*-deletion seems to be spreading. The test contained five sentences of the type in example (5) below.⁹

- (5) *Lehet, (hogy) én elmegyek addig fürdeni.*
 be.MOD.PRS.3SG (that) I go.PRS.1SG till.then bathe.INF
 ‘Perhaps I’ll have a bath till then.’

The corpus search in this case ran with different settings for several reasons. As the deletion of the complementizer seems to be an innovative feature, we searched only those subcorpora (spoken, personal, journalism) that were thought to be more likely to contain instances of the innovative form (i.e., where the complementizer is deleted). Besides, the size of the random samples was meant to be bigger, containing a thousand items from each subcorpus. However, as *lehet* can occur in different structures, and many

⁹ Many of the examples of *hogy*-deletion after *lehet* in our test stem from the collection of our colleague László Horváth, whom we sincerely thank for providing us with these data.

of these could not be filtered out automatically,¹⁰ after the manual analysis there was only a smaller subset of the original set of three thousand sentences left that could be classified with respect to the presence or absence of the complementizer. Finally, we ran a search with the same settings in the Transsylvanian subcorpus, as we had an impression that this phenomenon could be more frequent in certain eastern dialects of Hungarian. Figure 2 summarizes the results of the queries.

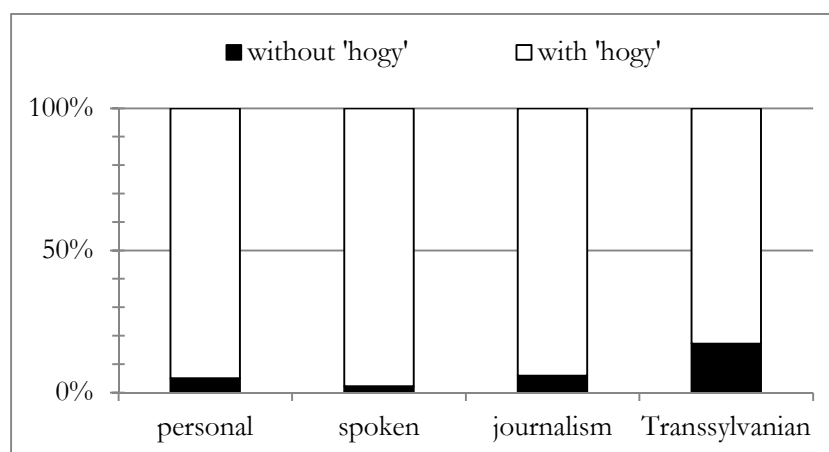


Figure 2: Sentences with *lehet* 'may be', with and without the subordinator *hogy*

On the basis of the data (see Figure 2), one could draw the following conclusions. First, although the innovative, unmarked structure is rare, it could be attested in all these subcorpora. Second, the impression that this is a regionally rooted phenomenon is confirmed by the data, as the unmarked form occurred most frequently in the Transsylvanian subcorpus. Finally, and not surprisingly, the frequency of the conjunction-less subordinate sentences with *lehet* as a matrix verb is low compared to the three verbs examined above (those in the case of which *hogy*-deletion is optional in MSH).

Naturally, the question arises how to analyze the innovative form structurally: is it the case that a further matrix verb joined the ranks of those verbs that allow complementizer-deletion, or do the innovative structures require a different structural analysis? In the latter case, the interpretation could be that this is an instantiation of grammaticalization across clauses, and the original matrix verb developed a new function, surfacing as a sentence adverbial, with its original subordinate clause in its scope. This type of change (during which biclausal structures become reanalyzed as monoclausal ones) keeps recurring with *hogy*-clauses throughout the history of Hungarian (Haader 2001), and, as a result, original matrix verbs are reanalyzed either as sentence adverbials or as conjunctions. In fact, the same verb *lehet* with the adverbial *jól* 'well' gave rise to the complementizer *jóllehet* 'although' ('It can well be that X [proposition]' → 'Although X'). Besides, this type of change (clauses of the type 'it may be that' giving rise to 'maybe'-adverbials) is common cross-linguistically as well (Beijering 2010). As it is not possible to

¹⁰ For instance, *lehet* can also function as a quasi-auxiliary, in which case the main verb is in the infinitive (e.g., *ezt meg lehet csinálni* 'this can be done'). These could be discarded automatically prior to generating the random samples by unselecting those cases in which there was an infinitive in the vicinity of the given verb. Still, numerous other types had to be removed during the manual analysis, as they were irrelevant from the point of the present analysis, like the frequent pattern in which *lehet* is a copula, occurring within a sentence containing a nominal or adjectival predicate (e.g., *okos lehet* '(s)he may be smart').

go into further details here (cf. Gugán 2015 for a more elaborate analysis), we can only suggest that many instances of the novel form are apparently ambiguous structurally (bi-clausal structure with omitted complementizer or monoclausal structure with a sentence adverbial), and it would be important to find the appropriate tests (potentially based on the prosodic pattern of the sentences) that could distinguish these two potential structures reliably.

The last set of sentences that focus on the presence or absence of *bogy* (phenomena 5, 6, and 7 in the test) do not contain matrix verbs at all. The striking feature of the given structure is that there is an adverbial that seems to govern a subordinate clause in this case, as in the examples (6)-(8).

- (6) *Természetesen, (bogy) neki van igazá.*
 Naturally (that) (s)he.GEN be.PRS.3SG truth.3SG
 ‘Naturally, (s)he is right.’ [lit. ‘(S)he has got his right.’]
- (7) *Valószínűleg, (bogy) fogalma sincs róla.*
 Perhaps (that) clue.3SG not.exist it.DEL.3SG
 ‘Perhaps (s)he does not even have a clue (of that).’
- (8) *Nyilván, (bogy) az én ebédemet ette meg.*
 Obviously (that) the my lunch.1SG.ACC eat.PST.3SG<3SG PV
 ‘Obviously, it was my lunch (s)he ate.’

This structure was first described in the seventies (see e.g., E. Abaffy 1976), but some highly sporadic instances can be attested already in Late Old Hungarian (Haader 2001). Owing to the truly unique nature of these structures, there were quite a few attempts to provide a structural analysis and/or an account of its development. A characteristic type of explanation is that this pattern is a result of the interference of two structures: an adjective that could subcategorize for a subordinate clause headed by *bogy* (9) and an adverbial that is the suffixed form of the same stem, but which could not have such a complement (10).

- (9) *Természetes, bogy neki van igazá.*
 Natural that (s)he.DAT be.PRS.3SG right.3SG
 ‘It’s natural that he is right.’
- (10) *Természetesen neki van igazá.*
 Naturally (s)he.DAT be.PRS.3SG right.3SG
 ‘Naturally, (s)he is right.’

Among others, E. Abaffy (1976) takes this position, but she also points out that the reanalysis of the given adverbials as adjectives (a recurrent type of change with certain types of adverbials throughout the history of Hungarian) could also motivate the appearance of such structures. Nemesi (2000) surveys the available analyses, and votes for a type in which there are in fact two clauses, and the adverbial modifies a deleted matrix verb and a deleted expletive pronoun (*Valószínűleg igaz az, bogy* [‘Probably ~~it is true~~ that’] → *Valószínűleg, bogy* [‘Probably that’]). Kenesei (2002) suggests that the functional element *bogy* in these cases is not an instantiation of the complementizer, and the structure itself is monoclausal, the adverbial being in the canonical position of sentence

adverbials. Finally, É. Kiss (2010) attributes the appearance of this structure to language contact with Rumanian. Her main argument is that this pattern is not quite compatible with the general structural characteristics of Hungarian, whereas several of the Romance languages use structures like this productively. Therefore, the structural description she offers relies on the structure hypothesized for the Rumanian pattern, meaning that this type of adverbials appears as the head of a special projection (SAP, Speech Act Projection) that subcategorizes for CPs. (For a more thorough review on the literature on this topic, see also the summary in É. Kiss 2010).

It is beyond the scope of the present paper to argue for any of these analyses. However, as a basis for comparing the test results, we also searched for the appearance of these structures in HNC. In this case, first we searched for all the instances three adverbials (*természetesen* ‘naturally’, *valószínűleg* ‘probably’, *nyilván* ‘obviously’). Then we searched specifically for those cases when these items are followed by *hogy*. These data were then analysed per subcorpora, during which the irrelevant data were sorted out.¹¹

Figure 3 shows the frequency of the adverbials occurring with *hogy* in each subcorpus. As the proportion of the adverbial with *hogy* is rather low, Figure 4 contains the occurrences of adverbial+*hogy* on a different scale, thus allowing the comparison of the subcorpora.

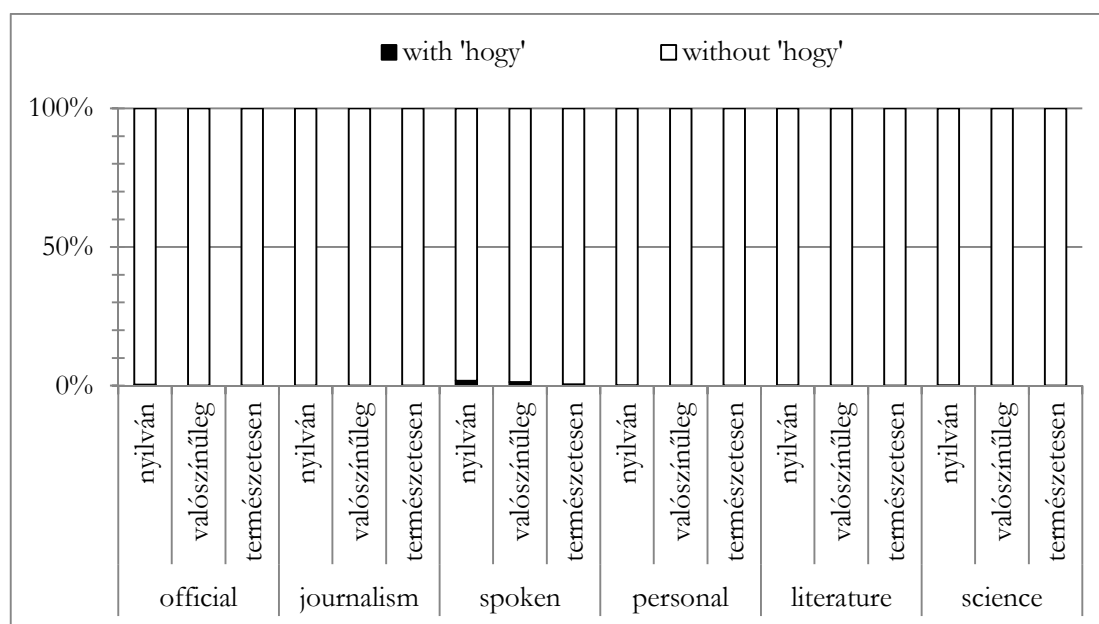


Figure 3: *Adverbials with and without hogy in the six subcorpora of HNC*

¹¹ I.e., those cases when the adverbial and the complementizer are adjacent, but the complementizer does not belong to the adverbial; this mostly meant hits in which the subordinate clause was an argument of the matrix verb, and the sentence adverbial modified that main clause as well, e.g.,

(i) *Úgy gondolom természetesen, hogy [...]*
 so think.1SG naturally that
 ‘Naturally, I think that [...]’

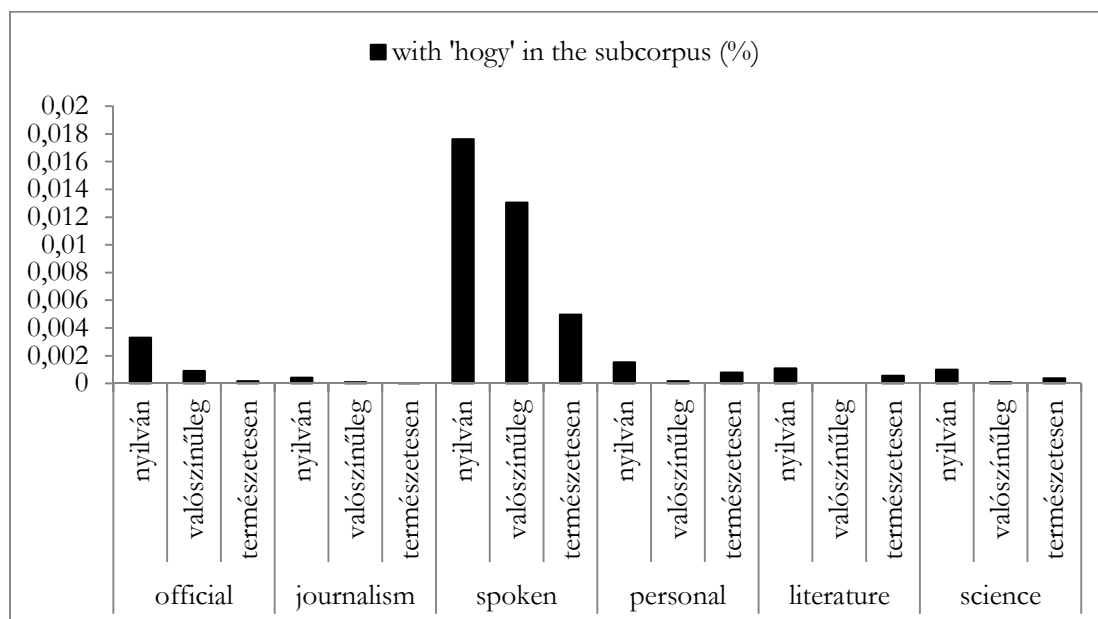


Figure 4: *Adverbials with hogy in the six subcorpora of HNC*

The most obvious remark to be made is that the adverbial+complementizer structure is quite rare: even in the spoken subcorpus, out of all occurrences of *nyilván* ‘obviously’, less than 2% of them co-occurs with the complementizer *hogy*. Here the following characteristics of the subcorpora should be recalled. The subcorpus “spoken” (where the adverbial+complementizer structure appears to occur with some frequency) contains utterances representing both formal and informal registers; the subcorpus called “personal” contains texts from forum comments from the Internet displaying a more informal register; and the subcorpus “official” is a mixture of read aloud speech and spontaneous speech. Consequently, the data suggest that the appearance of the structure in question is characteristic of spontaneous speech (a speaking style in which the speaker has less time to plan and monitor his/her speech production), and not necessarily characteristic of domains that can be characterized as more informal, allowing for innovative and/or potentially stigmatized variants (contra Kontra 2003).

4. Unmarked object (phenomenon 9)

In MSH, all objects are marked with the accusative suffix *-t* (11a-b).

- (11) a. *jön* *a* *hajó*
 come.PRS.3SG the ship
 ‘the ship is coming’
- b. *látom* *a* *hajót*
 see.PRS.3SG<1SG the ship.ACC
 ‘I see the ship’

There is one regular exception: its use is not obligatory with nouns marked with a Sg1 or Sg2 possessive suffix (12a-b).

- (12) a. *jön* *a* *fiam*
 come.PRS.3SG the son.1SG
 ‘my son is coming’
 b. *látom* *a* *fiam*
 see.PRS.3SG<1SG the son.1SG
 ‘I see my son’

As for the potential factors motivating the choice of the speakers to use the unmarked or the marked alternative, this seems to be totally optional grammatically, and neither of the forms is stigmatized or bound to registers. Diachronically, the unmarked accusative in this case is an archaic feature with parallels from e.g., Eastern Mansi (see Virtanen 2013); unfortunately, it is beyond the scope of this paper to survey the different historical explanations for this phenomenon (for details, see e.g., Korompay 1991, or É. Kiss 2014).

As there are no indications that register could be a factor concerning the choice between the two variants, the sample investigated from HNC was not filtered with respect to subcorpora.¹² We simply obtained random samples of nouns with Sg1 or Sg2 personal suffixes in the accusative, and manually classified the results according to the presence or absence of the accusative marker (besides, hits that were irrelevant for the present investigation were discarded). Figure 5 present the results of the corpus queries.

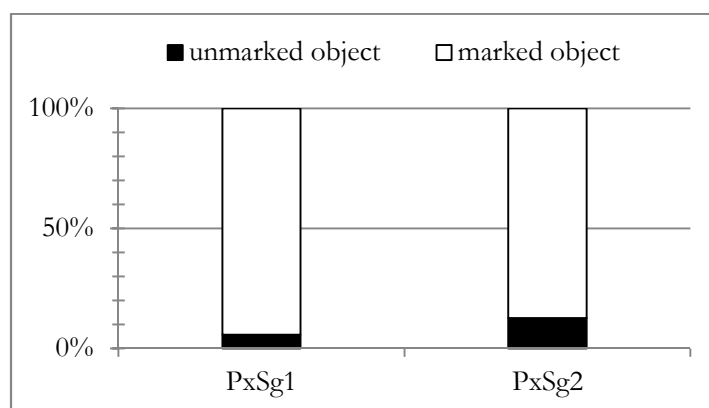


Figure 5: *Possessive-marked objects (with Sg1 or Sg2 personal suffixes) with and without the accusative marker*

It has to be mentioned that unmarked objects are in fact more frequent in the corpus than the data above would suggest, as the investigation of similar samples with the nominative form (i.e., nouns with Sg1 or Sg2 personal suffixes in the nominative) revealed that unmarked accusatives are in many case wrongly classified and thus appear in the lists containing nominative forms as well.¹³ However, even taking these findings

¹² The proportion of the two forms might still be different regionally, but this has not been investigated specifically.

¹³ We also checked two samples of nouns with Sg1 or Sg2 personal suffixes in the nominative. It turned out that 18 out of the 1000 nouns bearing a first person possessive marker, and 105 out of the 1000 nouns bearing a second person possessive marker were in fact unmarked accusatives. Besides, it is also telling that in a recently conducted test in which participants had to fill in slots in sentences under time pressure (Kalivoda 2016), the majority chose the possessive marked variant without the

into account, the unmarked, archaic pattern seems to be less frequent, but absolutely not marginal.

5 Unmarked plurality of the possessor on the possessee (phenomenon 11)

The expression of possession with third person plural possessors also displays variation in MSH.

First, the possessor (in all person-number variations) can be either in the nominative or the dative (13a-b). As opposed to the optionality of markedness described above in the case of object marking, variation here is not arbitrary: the different case assignments are due to the different structures of these two patterns (for a detailed structural description of the possessive constructions, see Szabolcsi & Laczkó 1992).

- (13) a. *Péter háza*
 Peter.NOM house.3SG
 ‘Peter’s house’
 b. *Péter-nek a háza*
 Peter-DAT the house.3SG
 ‘Peter’s house’

Second, the choice of the suffix marking third person plural possessor on the possessee depends on the grammatical category of the possessor. If it is encoded as a pronoun, the marker of the Pl3 possessor on the possessee (14b) is different from that of the Sg3 possessor (14a), whereas the pronoun itself is not marked for plurality. However, if the possessor is encoded as a noun, then the possessee is unmarked with respect to the plurality of the possessor (and the possessor bears a plural marker, as in (15a-b)).

- (14) a. *až ő háza*
 the he house-3SG
 ‘his house’
 b. *až ő háza-uk*
 the he house-3PL
 ‘their house’
 (15) a. *a fiú háza*
 the boy house-3SG
 ‘the boy’s house’
 b. *a fiúk háza*
 the boy-PL house-3SG
 ‘the boys’ house’

Naturally, the possessor could be in the dative case in the above examples (14)-(15) as well. From the point of view of the present discussion, the relevant pattern is when the plural third person possessor is encoded as a noun in the dative as in (16a-b).

accusative marker in the task when the context rendered it likely that the object would be a possessee as well.

- (16) a. *a fiú-k-nak a ház-a*
 the boy-PL-DAT the house-3SG
 ‘the boys’ house’
 b. *a fiú-k-nak a ház-uk*
 the boy-PL-DAT the house-3PL
 ‘the boys’ house’

Ultimately, this variation can also be explained on a diachronic basis. The dative suffix on the possessor (*-nak*) is itself an innovation compared to the unmarked possessor (i.e., unmarked dative), albeit an ancient one, as it emerged during Late Proto-Hungarian. Originally, the possessor bore only the plural marker *-k*, and the possessee was either unmarked or marked with respect to plurality (i.e., that of the possessor), so the pair that had coexisted first can be seen in (17a-b). The second member of this pair, i.e., the pattern showing number agreement (17b) is already obsolete in MSH.

- (17) a. *fiú-k ház-a*
 boy-PL house-3SG
 ‘the boys’ house’
 b. *fiú-k ház-uk*
 boy-PL house-3PL
 ‘the boys’ house’

There are good reasons to suppose that the non-agreeing form was the original pattern, and the two structures competed in Old Hungarian, but the non-agreeing form was prevalent in those texts that are closer to spoken language (Korompay 1991, 269). As for the dative-marked possessor in the same period, it was observed (Korompay 1992, 348) that whereas in structures where the possessee is not marked for the plurality of the possessor, the possessor can either be in the nominative or the dative (*fiú-k-0 ház-a* or *fiú-k-nak ház-a* were both frequently occurring types), in structures where the possessee is marked for plurality, the dative-marked form already seems to be obligatory (*fiú-k-nak ház-uk*). Therefore, the variation that is characteristic of MSH (as in 15) is already present in Old Hungarian.

Naturally, the coexistence of these forms must have been characteristic of the period between Old and Modern Hungarian as well, which is illustrated by the fact that, according to the survey of the handbook of normative linguistics (Grétsy & Kovalovszky 1980, 350), representatives of the language reform movement (taking place between 1790-1820 approximately) propagated the exclusive use of the agreeing form both with the nominative and with the dative possessor, meaning that they must have been aware of the presence of variation in agreement. It is also instructive to see how the authors of the handbook of normative linguistics interpret this variation concerning MSH: they note that the use of the non-agreeing form is spreading in MSH, but they also recommend the use of the plural-marked possessee with dative-marked possessors, especially in that case when the structure appears in an intransitive sentence expressing possession as in (18).

- (18) *A fiú-k-nak van ház-uk / (ház-a)*
 the boy-PL-DAT be.PRS.3SG house-3PL / (house-3SG)
 ‘The boys have a house.’

Therefore, as in this case it cannot be excluded that the non-agreeing form is associated with a specific register (and this explains why the authors of the handbook find it less adequate in MSH), the data collected from HNC are arranged according to subcorpora, and in this case again we chose to investigate those three subcorpora that are thought to be closer representatives of informal language use. The query in this case was set for nouns in the plural dative form, with the verbs *van* ‘there exists’ or *nincs* ‘there does not exist’ in their immediate environment. As hoped, the search results contained many instances of sentences expressing possession (‘He has got’ / ‘He does not have’), which is encoded in Hungarian with the possessor in the dative and the existential verb. (This way the data obtained through the corpus search can be easily compared to the test results, as the test also contained sentences expressing predicative possession). Figure 6 summarizes the results of this query. The random samples in this case contained 250 elements for each verb and each subcorpus, so altogether we analyzed 1,500 structures, but again, about a fifth of all data proved to be irrelevant from the point of view of the present discussion.

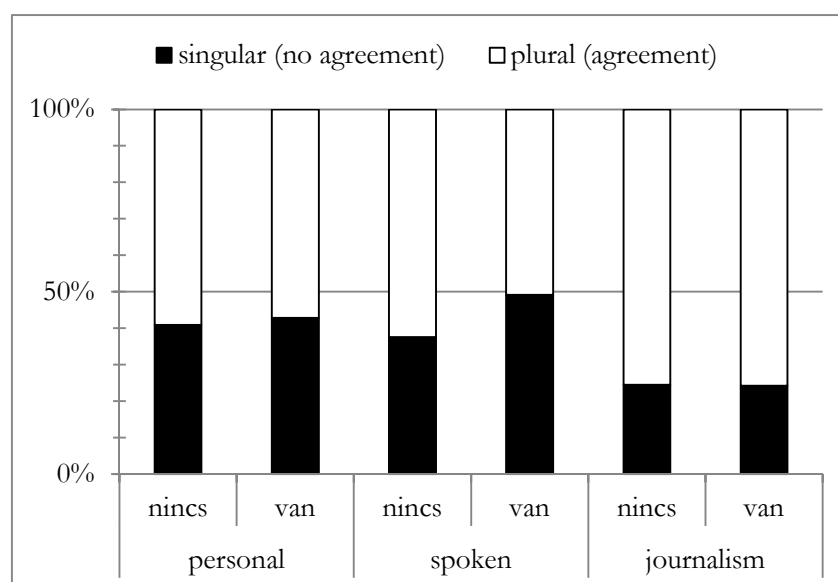


Figure 6: *Forms showing and not showing agreement in sentences expressing possession*

As it is apparent from Figure 6, although the agreeing form is in the majority with both verbs in all subcorpora, it could not be said that the non-agreeing form is infrequent; on the contrary, the speakers seem to use both forms widely.

6 Compounds of the type noun+verb (phenomenon 10)

There is a type of compounding in Hungarian in which the first stem of the compound is a noun, and the second is a verb, e.g., *apróhirdet* ‘to post small ads’, *bájcseveg* ‘to do small talk’, *agymos* ‘to brainwash’ (Kiefer 2000, 531). At first sight, these seem to consist of a verb and one of its arguments, the latter losing its appropriate morphological marker through the process of compounding, e.g., *város-t_{ACC} néz* ‘to go sightseeing’, lit. ‘watches the city’ → *városnéz*. However, this analysis would not be correct: these noun+verb type of compounds are backformations from pre-existing derivations in which the last stem is an action noun formed with the suffix *-ás/-és*, e.g., *apróhirdetés*, *bájcsevegés*, *agymosás*. Kiefer

(2000, 531) points out that sporadically, some compounds of this type may arise straight from a verbal phrase due to analogy, but this pattern of compounding is not (yet) productive in Hungarian. As these forms are morphologically transparent, all native speakers can interpret (and, theoretically, produce) the innovative, noun+verb compounds. However, the corpus query in this case could not help to survey the incidence of the novel forms, as such forms rarely occur in HNC: among the tested five compounds, only one (*ügyintéző* ‘to administrate’) was attested in the corpus as well. Therefore, this type of compound can by all means be assumed to be less frequent than the forms that can be the input of the word formation process (either the derived form with the action nominal, e.g., *ügyintézés* ‘administration’ or the noun + verb pattern, e.g., *ügyeket intéz*).¹⁴

Concluding this section, we have to emphasize that the aim of Experiment1 (see Section 3.2-3.3 in the article) was to single out linguistic variables that are subject to language change in MSH. For this purpose, we elicited grammaticality judgments concerning the above described variants, some of which we could have reliably labelled as “innovation” or “archaism” from a synchronic point of view prior to the grammaticality judgement test, and some of which we could only label as “less frequent” or “more frequent” on the basis of surveys in HNC. There were two reasons for collecting grammaticality judgments. First, we wanted to check whether any of the pairs that we referred to as “more” and “less” frequent corresponded to the opposition “conservative” and “innovative” from a synchronic point of view. Second, we wanted to check the validity of the previously assigned labels “conservative” and “innovative”. It is important to note here, however, that we did not expect all the tested variants to fit into the innovation-archaism opposition. On the contrary, we included some phenomena which were thought not to be representatives of this opposition, that is, which were thought to show stable variation (e.g., phenomena 9 and 10, that is, optional *bogy*-deletion and unmarked accusative). We hoped that those cases that seemed to be prototypical instances of stable variation would serve as a basis for comparison, facilitating the selection of true innovations.

References

- Bácskai-Atkári, Júlia and Éva Dékány. 2014. From non-finite to finite subordination: The history of embedded clauses. In Katalin É. Kiss (ed.), *The Evolution of Functional Left Peripheries in Hungarian Syntax*. 148–223. Oxford: Oxford University Press.
- Beijering, Karin. 2010. *The grammaticalization of Mainland Scandinavian MAYBE*. <https://bells.uib.no/bells/article/view/39> (2012. 10. 12.)
- E. Abaffy, Erzsébet. 1976. Valószínűleg, hogy...? *Magyar Nyelvőr* 100. 397–8.
- É. Kiss, Katalin. 2010. Valószínűleg, hogy román kontaktushatás. In Katalin É. Kiss & Attila Hegedűs (eds.), *Nyelvelmélet és kontaktológia*. 223–238. Piliscsaba: PPKE BTK.
- É. Kiss, Katalin. 2014. The evolution of functional left peripheries in the Hungarian sentence. In Katalin É. Kiss (ed.), *The Evolution of Functional Left Peripheries in Hungarian Syntax*. 9–55. Oxford: Oxford University Press.

¹⁴ A Google search in this case showed slightly different results: *tiszújít* [approximately: ‘elect new leaders’] seem to occur with some frequency in the language of the press mainly, and there were also hits for *magánrendel* ‘to do private consultations’, *osztálykirándul* ‘to go for a class outing’ and *vészfékhez* ‘to push the emergency brake’. Naturally, this result is not surprising, as Google searches the entire web, therefore, new patterns can be found more easily.

- Egedi, Barbara. 2014. The DP-cycle in Hungarian and the functional extension of the noun phrase. In Katalin É. Kiss (ed.), *The Evolution of Functional Left Peripheries in Hungarian Syntax*. 56–82. Oxford: Oxford University Press.
- Givón, Talmy. 2001. *Syntax. An Introduction. Volume 1*. Amsterdam: John Benjamins.
- Grétsy, László & Kovalovszky Miklós (eds.). 1980. *Nyelvművelő kézikönyv*. Budapest: Akadémiai Kiadó.
- Gugán, Katalin. 2015. Hol volt? Hol nem volt? A tagmondattörő grammatikalizációs folyamatokról a *lévén* és a *lebet* grammatikalizációja kapcsán. *Magyar Nyelv* 111. 38–53.
- Haader, Lea. 1991. Az összetett mondatok. In Benkő, Loránd (ed.), *A magyar nyelv történeti nyelvtana I*. 728–741. Budapest: Akadémiai Kiadó.
- Haader, Lea. 1995. Az összetett mondatok. In Benkő, Loránd (ed.), *A magyar nyelv történeti nyelvtana II/2*. 506–665. Budapest: Akadémiai Kiadó.
- Haader, Lea. 2001. Mikrodiakrónia és változásvizsgálat. *Magyar Nyelvőr* 125. 354–371.
- Harris, Alice C. & Lyle Campbell. 1995. *Historical syntax in cross linguistic perspective*. Cambridge: Cambridge University Press.
- Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.
- I. Gallasy, Magdolna. 1991. A névelő és a névelő-előzmény. In Benkő, Loránd (ed.), *A magyar nyelv történeti nyelvtana I*. 461–475. Budapest: Akadémiai Kiadó.
- I. Gallasy, Magdolna. 1992. A névelők. In Benkő, Loránd (ed.), *A magyar nyelv történeti nyelvtana II/1*. 716–771. Budapest: Akadémiai Kiadó.
- Kalivoda, Ágnes. 2016. *Szakít – félbe, szét, valakivel, valamit? Az igei komplexum feldolgozása produkciós feladatban*. Manuscript.
- Kenesei, István. 1992. Az alárendelő mondatok szerkezete. In Kiefer, Ferenc (ed.), *Strukturális magyar nyelvtan 1. Mondattan*. 529–714. Budapest: Akadémiai Kiadó.
- Kenesei, István. 2002. Hányféle igazság van? *Magyar Nyelv* 98. 39–49.
- Kiefer, Ferenc. 2000. A szóösszetétel. In: Kiefer, Ferenc (ed), *Strukturális magyar nyelvtan 3. Morfológia*. 519–568. Budapest: Akadémiai Kiadó.
- Kontra Miklós (ed.) 2003. *Nyelv és társadalom a rendszerváltás kori Magyarországon*. Budapest: Osiris.
- Korompay, Klára. 1991. A névszójelezés. A névszóragozás. In Benkő, Loránd (ed.), *A magyar nyelv történeti nyelvtana I*. 259–318. Budapest: Akadémiai Kiadó.
- Korompay, Klára. 1992. A névszójelezés. A névszóragozás. In Benkő, Loránd (ed.), *A magyar nyelv történeti nyelvtana II/1*. 321–410. Budapest: Akadémiai Kiadó.
- Nádasdy, Ádám. 2008. *Prédikál és szónokol*. Budapest: Magvető.
- Nemesi, Attila László. 2000. A *természetesen, hogy...* típusú szintaktikai szerkezetről. *Magyar Nyelvőr* 124. 430–442.
- Oravecz, Csaba, Tamás, Váradi & Bálint, Sass. 2014. The Hungarian Gigaword Corpus. In *Proceedings of LREC 2014*.
http://www.lrec-conf.org/proceedings/lrec2014/pdf/681_Paper.pdf
- Szabolcsi, Anna & Tibor Laczkó. 1992. A főnévi csoport szerkezete. In Kiefer, Ferenc (ed.), *Strukturális magyar nyelvtan 1. Mondattan*. 179–298. Budapest: Akadémiai Kiadó.
- Virtanen, Susanna. 2013. Contextual function of noun marking in the direct object–marking system of Eastern Mansi. In Csepregi, Márta, Kubinyi Kata & Jari Sivonen (eds), *Grammatika és kontextus: új szempontok az uráli nyelvek kutatásában III*. Uralisztikai Tanulmányok 20. 305–317. Budapest: ELTE Finnugor Tanszék.