Hungarian modal existential wh-constructions in written language use – diachronic perspectives

Introduction: In our presentation we examine Hungarian modal existential wh-constructions (MECs) as they appear in the written language of different periods. MECs have some universal properties, such as the presence of existential predicates (*van/nincs 'be', akad 'occour', lehet 'may be*[1]), a wh-pronoun, and a third element which is either a subjunctive verb or an infinitive. (1)

(1) *Nincs miről beszélni.* lit. 'There is nothing to speak about.'

The presence of these elements is compulsory (although the subordinate verb can be ellipted). Native speakers use MECs to express whether the availability of a proposition is possible or not[2]. Accordingly, MECs have a double core: the existential verb is the syntactic head and selects a non-finite or a subjunctive phrase, therefore MECs belong to the family of existential constructions. The other core is a semantic head, the non-finite or subjunctive verb which selects its own arguments (DP-s and PP-s), one of them realized as an interrogative pronoun. Considering that in Hungarian infinitives can be inflected, in MECs it is also possible to anchor the proposition to the participants of the discourse, as in related constructions containing an evaluative expression and an infinitive. • MECs have not been in the focus of many studies: the literature mostly deals with the formal attributes of them[1,3]. The aim of the current research is to provide supplementary corpus-based data to a better understanding of MECs. Our preliminary studies show that the construction has been productive and relatively frequent in language use. However, on the one hand, MEC has a less variable form than expected; on the other hand, it can occur in forms that are not mentioned in the literature. We analyze data on the basis of real language use, and not on possible, but rarely-used structures, also paying close attention to diachronic changes.

Research Questions: Which have been the most frequent MECs regarding the predicate, the wh-pronoun, and the infinitive form? • In what contexts have been inflected infinitives and subjunctive mood used? •How fixed has the construction been? •In what contexts and registers have been the most frequent MECs used? • What diachronic changes can be recognized in terms of MEC's idiomaticity?

Theoretical background: Since our aim is to analyze MECs in language use, we need a usage-based holistic framework, which treats all expressions (including idioms) equally and focuses on form and meaning simultaneously. In Construction Grammar (CxG) "[a]ny linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist." [4] We consider MEC to be a construction because its meaning cannot be clearly deduced from the meaning of its elements (especially because the pronoun does not have an interrogative function). Michaelis[5] claims that "every pattern of the language, from the fixed formulas to the fully productive phrase-structure rules" can be placed on a continuum of idiomaticity. Based on our data MECs can be divided into subgroups in terms of idiomaticity: they most frequently appear as partially fixed idiomatic expressions, such as *nincs mit tenni* lit. *'there is nothing to do'*. Some expressions, such as *kezd vele valamit* lit. 'start something with it' or *tejbe aprít* lit. 'chop into milk' occur almost exclusively in MECs. However, other MECs, such as *nincs miért mosolyognom az égre* lit. 'there is nothing why to smile on the sky' show that the MEC-pattern can be productively used to express new meanings in new contexts.

Methods: Data has been collected in the Old Hungarian Corpus (OHC)[7], in The Old and Middle Hungarian corpus of informal language use (OMHC)[8], in the Hungarian Historical Corpus (HHC)[9], and in the Hungarian Gigaword Corpus (HGC)[10]. Corpus-based data was needed since previous empirical studies were not able to give clear answers to the listed questions. Due to the large amount of data (having more than 27 thousand items in HGC), the pilot research was carried out on a manually unsorted sample, therefore at this point, we are only able to show tendencies in the usage of MECs. It must be pointed out that finding the relevant appropriate data on different query interfaces is already an achievement, because of the partially free structure of MECs. Concerning the contexts and registers in which MECs can appear we need further investigation.

Results • Frequent forms: As opposed to Szlávich[6], in OHC we got a list of 36 examples, 33 of them were negative statements. The main predicate was the substantive verb, *tud* 'can' appeared 7 times. Generally, MECs were used to express the lack of basic human needs (*nincs mit egyenek* lit. 'for them having nothing to eat'). In OMHC 162 MECs could be found, half of which was introduced by the verb *tud*; the other part stood with some form of the substantive verb. Most examples were negated in this sample too. Since this corpus consists of written texts, partly of personal letters, MECs were embedded in new contexts: e.g. to express the inability to write about a topic (2).

(2) *A doktor felől most sem tudok kegyelmednek mit írnom.* lit. 'about the doctor I cannot write anything for you now either'

In the HHC more than 1800 exemplars are found, in HGC more than 27 000. Also in the HHC and HGC, MECs are mostly introduced by the negated substantive verb (in the HHC and in the HGC the verb *tud* occurs in 15-20% of the total data), whereas some predicates mentioned by Lipták[1], such as *akad*, or *bír*[2] hardly ever appear.

Infinitives and subjunctives: In OHC the rate between the infinitive and subjunctive forms was different from today: only 53% of the data contained an infinitive, and inflected forms were more common than non-inflected forms. The choice between the inflected and non-inflected variants seems to be arbitrary. The dative-marked logical subject occurs rarely, if it appears, the infinitive is non-inflected (3).

(3) *emberfiának kedig nincs hova fejét hajtani.* lit. 'one does not have where to lay his head' MunchK - 66vb/Lk_9/58 - 3/276860

The query brought a remarkable finding to the light: in contrast to synchronic variation, in OHC and OMHC an inflected infinitive was also possible besides the (already inflected) main predicate *tud.* (2) (4).

(4) *megkérdezvén őtet tudjak mit írnom* lit. 'asking him what I can write' JordK - 791/ApCsel_25/26 - 5/201731

In OHC, 15 examples contained a subjunctive verb form, and one example contained a conditional, which can be regarded as a negative transfer in translation. Subjunctive variants occasionally contained a relative-like pronoun[1]. In OMHC: Out of the 162 examples, there was only one which had a subjunctive as the subordinate verb, which is interesting considering the larger number of such examples in the OHC. In the constructions having the substantive verb as the main predicate, 45 MECs were inflected and 36 MECs stood with a bare infinitive. Szlávich[6] claims that only 7,7% of the data contains a subjunctive form in HHC, and 2,19% of the data contains a subjunctive form in HGC, which is supported by our data. In HHC 65% of the strictly bound constructions had uninflected infinitive, 15% of the infinitives stood with the 3. person singular inflection, whereas 10% was in the first person singular. The majority of the infinitival constructions are not inflected today.

In respect to **word order**, in OHC most of the examples are strictly bound, however — as opposed to Lipták[1] — discontinuous forms are also possible: adverbial elements could stand between the verb and the pronoun, nominal arguments of the infinitive could stand between the pronoun and the non-indicative form (5). Sluicing, that is IP-ellipsis was also possible (6).

(5) *nincs annak utána továbbá mit tegyenek.* lit. 'there is not to do after that' MunchK - 70ra/Lk_12/4 - 3/279076

(6) *láték tűzlángot (…), de az egyebekről nem tudok mit* . lit. 'then I saw flames (…) but about others I can't' JokK - 45/17 - 1/169881

In OMHC: most part of the sample shows the strict verb-pronoun-infinitive order, however, the intermediate positions can be expanded with one or two words. Although there are some examples having a word in between the pronoun and the infinitive, there are more examples that have an adverbial or a noun phrase between the main verb and the pronoun (7).

(7) *nem tud nyomorult Leányával mit tenni* lit.'(s)he cannot do anything with her/his miserable daughter'. 1062843

In HHC, 80% of the data showed the strict verb — wh-word — infinitive word order, in HGC 75% of the cases are fixed. Those constructions where the infinitive was the initial element appeared in less than 1% of the sample in HGC. 10% of the sample had examples in which there was another word between the main verb and the wh-pronoun. These elements are typically particles and adverbs (*is* 'also', *még* 'even*). IN HHC,a frequently appearing example is the expression *nem tud(ok) vele mit kezdeni* (lit. 'I can't do anything with/about it'). There were only a few examples where this position was extended by more words, even in HGC.

Idioms: Contrary to earlier periods, in HHC and HGC, the idiomatic expression *nem volt/nincs mit tenni* lit. 'there is/was nothing to do' was noticeably the most frequent construction. The expression is typically used when the speaker wants to convey that s/he has no opportunity to act, or it is not worth making any effort and has already accepted this state. In addition, speakers often express their regret, at the same time indicating that they have no influence on the course of events. Knowing the context, therefore, is essential in the case of MECs, it is not sufficient to acknowledge only the meaning of the elements making up the construction (9).

(9) *Ha nincs mit tenni, fogadjuk el, és tanuljunk belőle.* lit. 'If there is nothing to do, let's accept it and learn from it.' #1192198523

Conclusion: MEC-frame is relatively productive, and strictly bound in written language use, even in Old Hungarian, although some elements could and can appear between the obligatory elements (mostly adverbs). However, there is no sign of complex structures occupying these positions. ● While the ratio between the subjunctive and infinitive forms was equal in Old Hungarian, in current Hungarian the vast majority of the examples stands with an infinitive. ● In New Hungarian and in the current language state there are fewer inflected infinitives than in Old Hungarian, the dative-marked subject appears rarely. ● In Old and Middle Hungarian, the infinitives could be inflected even if they follow the main predicate *tud.* ● Even speakers of Old Hungarian used MECs to express the possible (in)availability of a proposition. ● This construction has been used most frequently in a negative form. ● Certain structures have turned idiomatic over time — these are the most frequent examples in today's corpora.

Bibliography: 1. Lipták A. 2003. *Hungarian modal existential wh-constructions.* University of Leiden. ● 2. Šimík, R. 2011. *Modal existential wh-constructions.* LOT, Utrecht ● 3. Surányi B. 2005. *Indefinite wh-pronouns and the modal existential wh-constructions in Hungarian.* Abstract from ICSH7. ● 4. Goldberg, A. 2006. *Constructions at work: The nature of generalization in language.* Oxford: Oxford University Press, 5. ●5. Michaelis, L. 2017. *Meanings of Constructions.* ● 6. Szlávich E. 2018. Nincs mit mondjak - egy konstrukció elemzése.*Ösvények I:* 54-71. ● 7. OHC: Simon E., Sass B. 2012. Nyelvtechnológia és kulturális örökség, avagy korpuszépítés ómagyar kódexekből. *Általános Nyelvészeti Tanulmányok.*(XXIV): 243-264. ● 8. OMHC: Novak A. et al. 2018. Creation of an annotated corpus of Old and Middle Hungarian court records and private correspondence. *Language Resources and Evaluation* 52:1–28. ● 9. Sass, B. 2016. A kibővített Magyar történeti szövegtár új keresőfelülete. *A nyelvtörténeti kutatások újabb eredményei IX.* 2016. április 27., Szeged ●10. Oravecz, Cs. at al. 2014. *The Hungarian Gigaword Corpus.* In: Proceedings of LREC 2014.