



## The form and function of interrogatives: A corpus-based study of Hungarian questions

*Alexandra Markó*

### Introduction

Questions are ubiquitous in everyday social interaction, therefore they have been investigated at every level of linguistic analysis for decades (Freed 1994). Research on the prosody, syntax, logic, semantics, pragmatics, psycholinguistic nature and conversational usefulness of questions can be found in a wide range of papers and books (see eg Austin 1962, Searle 1969, Schegloff & Sacks 1973, Hymes 1974, Fishman 1978, Kiefer 1983, Vion & Colas 2006). Although there is a general agreement that questions have an identifiable syntactic form, characteristic prosody, a semantic or propositional content which is separate from their pragmatic and social function and separate also from their various roles in conversational interaction, there is considerable disagreement about how questions should be defined, classified and analysed (for a review, see Freed 1994).

The interactional approach suggests that the analysis of questions should take into account the context in which they occur and the local situated interests of discourse participants, however the recognisability of questioning can be derived from purely formal linguistic features as well (eg syntax or intonation) (Rossano 2010). Freed (1994) posits a taxonomy illustrating how questions vary along an information continuum. Sixteen different functional categories are established, from public information (which are about the external world and request new factual information) to reported speech. In between, one can find such functions as social invitation (eg *Would you like some. . . ?*), clarification of information, repetition of information (eg *Pardon?*), phatic information (eg *You know what I mean?*), didactic, rhetorical function and so on. The 16 continual types of Freed (1994) correspond to four functional classes recognized by the previous literature, see for example Kearsley's overview of questions (1976).

The function and the form of questions show interrelations (see Freed 1994 among others).

Based on their logical semantic structure, three main question types are differentiated: (1) content questions (in English, they are often referred to as *wh-questions*), (2) polar questions (also known as *yes–no questions*) and (3) alternative questions.

Content questions are interrogative sentences that contain a question word such as *who* or *where*, and the focus of the expected answer belongs to the same ontological category as the question word (eg a person reference in response to ‘who’, a spatial reference in response to ‘where’). Open questions are considered a special type or offshoot of *wh-questions*, with the same syntactic structure but different semantic features: they require longer exposition (not just one word) as an answer (eg *How can we solve this problem?*) (Kiefer 2000, Kugler 2000). In Hungarian prosody, both *wh-questions* proper and open questions feature a front falling contour (Deme 1962, Varga 2002a).

Polar questions require affirmation/confirmation or disconfirmation. Alternative question is considered the third main category (beside content and polar question) by several authors (eg Stivers 2010, Rossano 2010), while other scholars classify it as a subtype of polar question (Freed 1994, Kiefer 2000, Kugler 2000). In Hungarian, polar and alternative questions have quasi-identical syntactic structure (Kugler 2000), however in terms of prosody they are rather different (see Olaszy 2002, Varga 2002a).

Questions can be marked formally by lexical, morphological, syntactic or prosodic devices appropriate to the given language, however questions are sometimes produced without such marking and conversely, not all interrogative sentences which are characterized by the formal features of the category can be evaluated as “real questions” (requiring an answer) from an interactional perspective.

Hungarian content questions are lexically marked (question word). Polar questions can be marked by a question particle or a special (final rise-fall) intonation contour (the distribution of morpho-lexical and prosodic markers showing complementary distribution), as well as with a question tag. The obligatory marker of alternative questions is *vagy* ‘or’, and a complex prosodic structure (eg high monotone + fall, see Varga 2002a).

In 2010, a special issue of *Journal of Pragmatics* was dedicated to the pragmatics of questions and their responses in 10 languages (from Europe, the USA, Southeast Asia, Mexico, Namibia, and Papua New Guinea). Although the analysis was predominantly pragmatic in orientation, several other factors were also necessarily taken into consideration, including log-

ical semantic structure, morphosyntactic and prosodic characteristics. In the survey, approximately 350 question-answer pairs taken from informal conversations were analysed in each language.

In the Hungarian literature, there is an abundance of theoretical analyses of interrogatives in terms of semantics, syntax, pragmatics or prosody (see, eg Deme 1962, Fónagy 1998, Fábri 1981, Kiefer 2000, Gósy & Terken 1994, Olaszy 2002, Varga 2002b). Nevertheless, research on questions as occurring in spontaneous speech is somewhat underrepresented. Fónagy & Magdics (1963, 1967) based their findings primarily on the intonation of questions in read and rehearsed by actors. Although spontaneous realizations are also mentioned in their work, these are sporadic, lacking a systematic analysis. Pragmatic studies on interrogatives in Hungarian have mostly focused on (semi)institutional debates (eg Schirm 2007).

In a previous study, I recorded a conversation involving four 21-year-old native speakers of Standard Hungarian (2 female and 2 male subjects). The length of the recording was nearly two hours, in which (after the exclusion of noisy occurrences affected by overlapped speech, laughter, etc) 199 questions were analysed in terms of the phonetic characteristics of intonation (Markó 2007). It was found that the intonation structure of spontaneous questions is much more diverse than that of read or rehearsed utterances as described in the literature. The conversation provided samples of numerous alternative realizations of phonological forms which had not been documented in earlier studies, and several questions arose with respect to the identification of the function of interrogatives.

In the present research, I have analysed informal conversations focusing on functional and logical semantic types as well as prosodic realizations. According to the hypotheses, (1) the assignment of conversational questions into logical semantic types depends on the function of the utterances involved (whether or not they require an answer), and (2) the prosodic realization of interrogative sentences depends on their function as well. Specifically, the occurrence of irregular or unexpected contours is more frequent among “non-real” interrogatives (which do not call for an answer) than among “real” ones.

## **Material and method**

The material of the analysis reported here was selected from the BEA Hungarian speech database (“BEA” stands for *beszélt nyelvi adatbázis* ‘spoken language data base’), recorded at the Research Institute for Linguistics of the Hungarian Academy of Sciences under constant circumstances in an anechoic chamber ([www.nytud.hu/dbases/bea](http://www.nytud.hu/dbases/bea); for detailed technical pa-

rameters, see Gósy 2012). BEA's material consists of several speech samples from various types of spontaneous speech, interviews, repetitions of stimuli, reading aloud, and conversations. The subjects are monolingual native speakers from Budapest aged between 20 and 90 years.

For the present research, 30 conversations were selected from BEA. In the conversation module of the database, there are three participants: the subject, the interviewer, and a third person. The topics vary, but invariably concern everyday life. Some conversation topics are: Easter, marriage vs cohabitation, secondary school final exams, summer holidays, school violence, keeping pets in an apartment, legalization of light drugs, theatrical life, students' rights, women's careers, the value of a university degree, etc. Topics for the conversation module are selected by the interviewer in accordance with the subject's age, job, and area of interest (based on the interview module).

The 15 female and 15 male subjects were selected on the basis of age from three age groups, see table 1. Since the distribution of the informants' age is not completely balanced in the BEA database, the range of age is not necessarily identical in the age groups of female and male subjects.

table 1: *The distribution of subjects in terms of sex and age (in years)*

	Female subjects		Male subjects	
	Range	Average	Range	Average
Young	21–24	22.2	21–24	22.0
Middle-aged	39–46	43.6	39–42	40.0
Elderly	54–60	57.4	57–66	61.0

The total duration of the 30 conversations exceeds 10 hours. Of this, the subjects' speech amounts to 290.5 minutes (ie approximately 5 hours). The duration of conversation recordings ranges between 6 minutes 42 seconds and 73 minutes 26 seconds, the average being 20 minutes 50 seconds. The subjects' speaking time was measured between 2 minutes 38 seconds and 45 minutes 12 seconds, with an average of 9 minutes 42 seconds.

Questions were identified in each recording. Because of the special background of BEA recordings (the interviewer and the third participant in all 30 conversations are from the same 4 persons developing the database, so their speech is overrepresented), only the questions asked by the subjects were taken into account in the analysis. Those which were suitable for a complex analysis (ie those not affected by overlapping speech, laughter or any other noise) were categorized according to the following criteria:

- (1) logical semantic structure (polar, content or alternative);
- (2) function (whether the question requires an answer or it has a different function);
- (3) prosody (phonologically regular or not).

The context that preceded and followed each question was considered in the functional classification of interrogative sentences. With respect to logical semantic structure, first the total corpus of questions was analysed. In a next step, the functional and the formal approach were combined, and a logical semantic classification was carried out separately for the “real” questions (which require an answer) and the “non-real” ones. In the functional analysis, the method introduced by Stivers & Enfield (2010) was followed.

## Results

In the approximately 5 hours of the analyzed speech material of 30 subjects, a total of 134 (not noisy) question tokens were labelled. On average, this means that the subjects uttered an interrogative in every 2.2 minutes. Naturally, there were some subjects (9 in total) who did not ask a question at all in the entire conversation, while a young man used 31 interrogative sentences (in 90 minutes of speech), and 20 interrogatives were found in an older man’s speech material (which lasted more than 30 minutes). As a general rule, no correlation can be found between speech duration and the number of interrogatives. For example, some speakers failed to use any questions despite talking for more than 30 minutes. The average number of interrogatives was 4.5 per person.

Regarding the logical semantic characteristics of the analyzed interrogative sentences, the two main types are represented nearly equally. The ratio of content questions is 45.52% (61 occurrences), the ratio of yes–no questions is 52.24% (70 tokens), and 3 alternative questions also occurred (2.24%).

The normative implications of questions (ie the occurrence of a response) are interactional in nature and pertain to the sequential organization of social action and its accountability (Rossano 2010). From this interactional point of view, a rather numerous group of interrogatives was evaluated as “non-real” in the present material. In these cases, the interlocutor does not expect an answer. The number of these interrogatives was 56, accounting for 42.42% of the total amount of tokens. This ratio seems to be relatively large, therefore I compared it to the data reported in the literature. A corpus of American English questions (1275 tokens) collected from dyadic conversations was analysed by Freed (1994). In her data, so-called relational and expressive questions have a share of 51.69%. In an-

other study of American English, informal social interactions were video-recorded in familiar settings with 3 to 5 speakers. In this corpus of 350 interrogatives, 6.29% of the questions were “non-functional”, in the sense of apparently not carrying any expectation of a response (Stivers 2010). Japanese interactions (informal, spontaneous conversations among 3–4 adult friends and/or family members) examined by the same project contained 8.57% “non-functional” questions in a sample of 350 (Hayashi 2010). These data suggest that the ratio of questions not requiring an answer can be influenced by the interaction’s style in terms of the formal-informal scale, and other factors such as the topic of the conversation.

In the corpus under investigation, almost half of the “non-real” occurrences (49.09%, 27 tokens) belonged to reported speech, where the question contained information from an entirely different (past or hypothetical) speech event. 19 questions (33.93% of the “non-real” category) had a rhetorical function, 7 tokens (12.50%) were realized as phatic parenthesis (eg *tudod?* ‘you know?’). 3 questions (5.36%) expressed the uncertainty of the speaker, where the speaker was aware that the hearer did not know the answer for the question, but (s)he was not confident of the content of her/his utterance (see B’s answer in the following sequence, uttered with the intonation contour of yes–no questions: A: *Hány óra volt az út?* ‘How many hours did the trip take?’ B: *Hát olyan tizenkettő?* ‘Well, about twelve?’)

In 78 cases (57.58% of the total corpus), the question was followed by a (verbal or non-verbal) answer. The logical semantic classes of these “real” interrogatives show completely different ratios compared to the whole material, and compared to “non-real” questions, too (figure 1). The ratio of polar questions is overrepresented among the “real” questions at 65.38% (51 tokens), with content question only accounting for 32.05% (25 tokens), and alternative questions 2.56% (2 tokens). The 56 occurrences of “non-real” questions include 33.93% polar (19 tokens), 64.29% content (36 tokens) and 1.79% alternative questions (1 token).

For the sake of comparison, it is worth noting the data of a similar analysis of questions in other languages published in the special issue of *Journal of Pragmatics* (2010). Even though the cited studies only used data from maximally informal social interaction in familiar settings between people who knew each other well (Enfield et al. 2010), and the recordings of the Hungarian BEA database are more or less staged, the comparison of the raw results can be fruitful (figure 2). The distribution of logical semantic types in “real” questions is quite similar in the conversations of various languages, which may suggest that this distribution is universal.

figure 1: The logical semantic distribution of “real” and “non-real” questions in the corpus

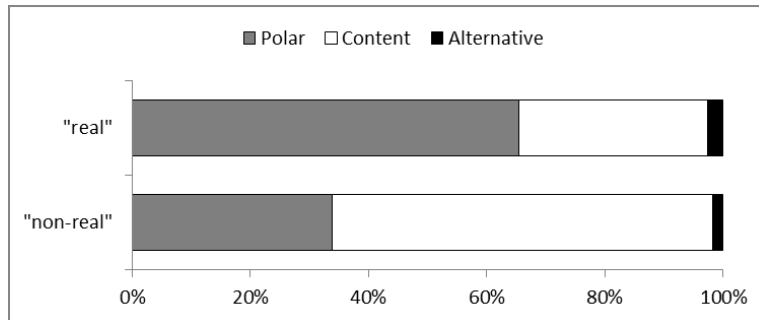
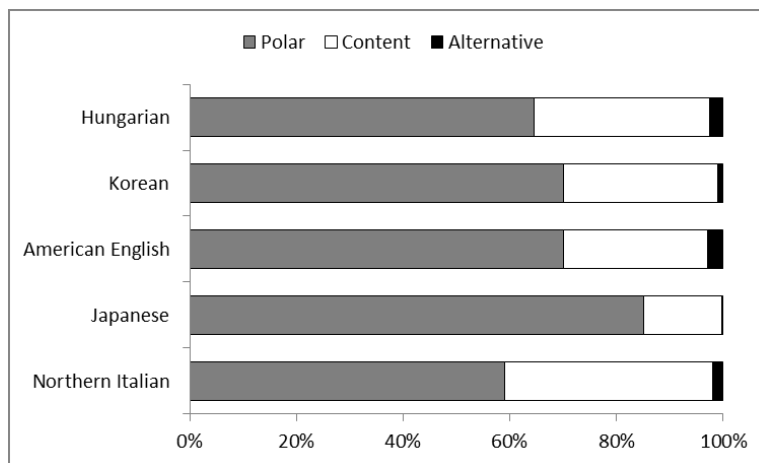


figure 2: Distribution of polar, content and alternative (“real”) questions in Hungarian (present research), Korean (Yoon 2010), American English (Stivers 2010), Japanese (Hayashi 2010) and Northern Italian (Rossano 2010) conversations



The prosodic analysis was only carried out on 125 interrogatives, since 5 content questions were realized with glottalization (4) or breathy voice (1), and 4 further tokens (3 polar and 1 content) were excluded because of their attitudinal characteristics (expression of disappointment). First the frequency ratio of regular prosodic structure was defined both for the overall data sample and specifically for “real” and “non-real” questions as well as for each logical semantic type. It should be noted that when the prosody differed in an expected manner because of a special function of the question, occurrences were regarded as regular. For example, echo questions are identical with content questions in their syntactic structure, but their intonation contour is characterized by a final fall-rise. With respect to Hun-

garian content questions, there is an ongoing debate on a relatively frequent intonation variant, namely the fall-rise, the question being whether it can be accepted as a regular form or not (for a review, see, eg Gósy 1993, Olaszy 2002). Since in the phonological description of Hungarian intonation (Varga 2002a), it is considered a variant of the basic form (with the special function of expressing conflict), in the present analysis I also adhere to this classification.

The numbers and ratios can be seen in table 2. The data do not show significant differences between “real” and “non-real” questions in terms of regularity of prosody, but polar “real” questions tend to be somewhat more regular than “non-real” ones. In the case of “non-real” questions, unexpected prosodic variations are more frequent among polar questions.

table 2: Ratio of regular prosodic structure in the various functional and formal types of questions (A = number of tokens, B = number (and ratio) of tokens with regular prosody)

		Content	Polar	Alternative	Total
“Real” questions	A	24	49	2	75
	B	19 (79.17%)	39 (79.59%)	1 (50.00%)	59 (78.67%)
“Non-real” questions	A	31	18	1	50
	B	23 (74.19%)	12 (66.67%)	1 (100.00%)	36 (72.00%)
All questions	A	55	67	3	125
	B	42 (76.36%)	51 (76.12%)	2 (66.67%)	96 (76.80%)

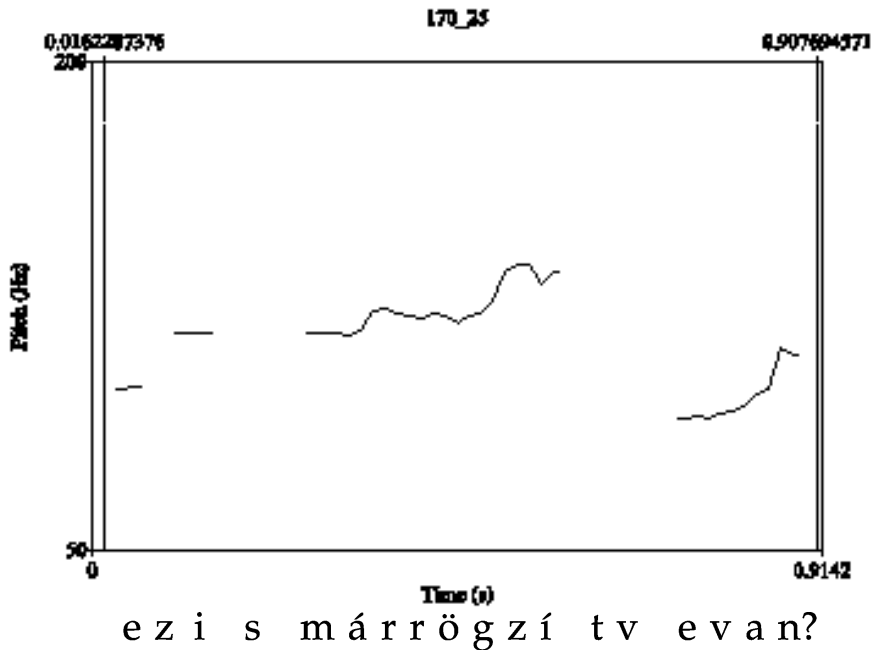
As was mentioned above, 3 questions were found in the corpus which expressed the uncertainty of the speaker. In terms of formal characteristics, one of them belonged to polar questions, realized with (regular) final rise-fall, while two of them were content questions with the intonation contour of echo questions, that is the same final rise-fall. Furthermore, 3 rhetorical (therefore formally content) questions were documented with final rise-fall as well (eg *Miről beszélünk?* ‘What we are talking about?’). These results suggest that, reflecting a special intention on behalf of the speaker, content questions can be realized with the intonation contour of polar questions, so this possibility is not limited to echo or repetitive questions.

The polar questions realized with an intonation contour different from the expected final rise-fall showed mainly the same variants as in Markó (2007). In one third of the cases, the rise-fall contour was followed by another rise (or after the rise the  $f_0$  stayed high). In a similar proportion of tokens, instead of a rise, a high monotone plateau was followed by a fall (which was not motivated by a peculiar attitude). In several other cases,



the position of the f0-peak was shifted from the penultimate syllable to a previous one, and this was also occasionally combined with a final rise (see figure 3).

figure 3: Pitch of the polar question *Ez is már rögzítve van?* 'Is this already being recorded?' with two peaks

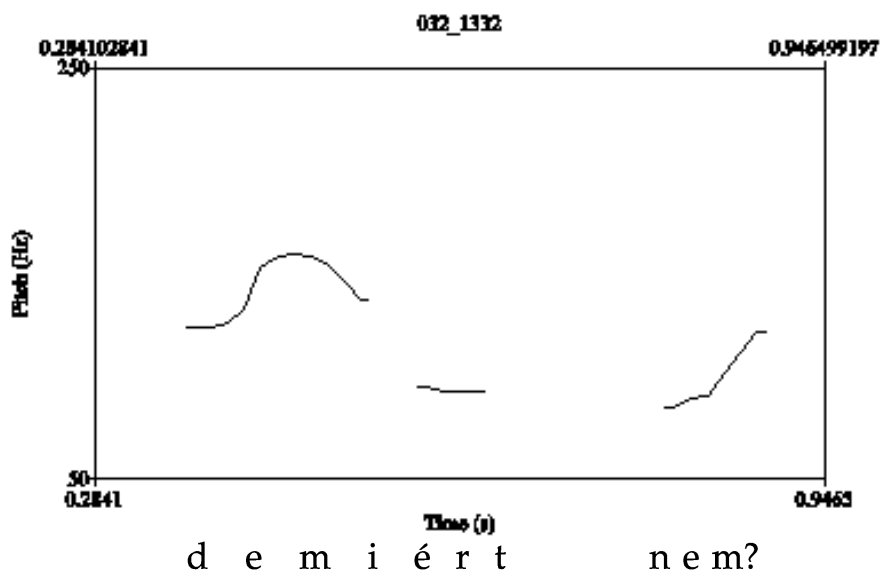


Although this subtype is considered to be frequent, not more than 8 of the content questions ended by a final rise. Despite what conventional wisdom and some remarks in the literature might suggest, this form was documented in men's speech with only one exception (see figure 4).

## Conclusions

The analysis reported here focused on the functional and logical semantic types of interrogative sentences and their prosodic realizations in 30 conversations of the BEA speech database. Based on conversational data from five languages, it was found that the distribution of logical semantic types does not depend on the language, with polar questions rated the most frequent by all of the cited studies. At the same time, there may be correlations between the interaction's level of formality and the distribution of polar and content questions (Kearsley 1976). In the cited studies as

figure 4: Pitch of the content question *De miért nem?* 'But why not?' with final rise from a male speaker's sample (the break of continuity in the  $f_0$ -curve between *i* and *é* is caused by irregularity of voice)



well as in the present research, the analysed conversations were more or less informal, which may partially account for the similar results.

It was presumed that the distribution of logical semantic types of questions in conversations would depend on the function of the interrogative sentence (whether it required an answer or not). This hypothesis has been confirmed: the ratio of content and polar questions is exactly the opposite among "real" and "non-real" questions. This result can be the consequence of pragmatic characteristics of the two groups of questions. Probably in a more informal situation, in which the social distance between the interlocutors is low, "real" questions are more frequent, and—as we have seen—among them polar questions are predominant.

The prosodic realization of interrogative sentences was assumed to depend on their function. It was found that the occurrences of irregular or unexpected contours were more frequent among "non-real" questions than among "real" ones, but this difference was considerable only in the group of polar questions. This result may be explained by the fact that in the case of "real" polar questions (lacking any other marker) the identifiable intonation contour is a very important cue for perception.

The analysis of spontaneous Hungarian questions suggests that in approximately 20–30% of the occurrences (after excluding irregular tokens,

eg those affected by speaker emotion), the intonation of questions does not support the findings of the literature. Some alternative realizations can probably be explained by pragmatic factors which were not taken into account in previous studies to a sufficient degree. The results raise the question whether the unexpected prosodic realizations are indeed irregular, or they should be evaluated as variants of the phonologically defined contour. There is reason to believe that just as the fall-rise contour of content questions has been accepted as a regular variant in the wake of recent research, the further analysis of spontaneous speech and various realizations of Hungarian questions may render our knowledge of interrogative functions and their mapping to intonation forms more precise.

#### REFERENCES

- Austin, John L. 1962. *How to Do Things with Words*. Cambridge, MA: Harvard University Press.
- Deme, László. 1962. Hangsúly, szórend, hanglejtés, szünet. [Stress, word order, intonation, pause]. In: József Tompa (ed.), *A mai magyar nyelv rendszere. Leíró nyelvtan II*. [The system of present-day Hungarian. Descriptive grammar, vol. 2.]. Budapest: Akadémiai Kiadó. 457–522.
- Enfield, N. J., Tanya Stivers, and Stephen C. Levinson. 2010. Question–response sequences in conversation across ten languages: An introduction. *Journal of Pragmatics* 42: 2615–2619.
- Fábricz, Károly. 1981. Az *-e* kérdő partikula. [The *-e* question particle]. *Magyar Nyelvőr* 105: 447–451.
- Fishman, Pamela. 1978. Interaction: The work women do. *Social Problems* 25: 397–406.
- Fónagy, Iván. 1998. Intonation in Hungarian. In: Daniel Hirst and Albert di Cristo (eds.), *Intonation Systems. A Survey of Twenty Languages*. Cambridge: Cambridge University Press. 328–344.
- Fónagy, Iván and Klára Magdics. 1963. A kérdő mondatok dallamáról. [On the intonation of interrogative sentences]. *Nyelvtudományi Értekezések* 40: 89–106.
- Fónagy, Iván and Klára Magdics. 1967. *A magyar beszéd dallama*. [The intonation of Hungarian speech]. Budapest: Akadémiai Kiadó.
- Freed, Alice F. 1994. The form and function of questions in informal dyadic conversation. *Journal of Pragmatics* 21: 621–644.
- Gósy, Mária. 1993. A kiegészítendő kérdés dallamváltozása. [Change of intonation in content questions]. *Magyar Nyelvőr* 117: 443–447.
- Gósy, Mária. 2012. BEA — A multifunctional Hungarian spoken language database. *The Phonetician* 105/106: 50–61.
- Gósy, Mária and Jacques Terken. 1994. Question marking in Hungarian: timing and height of pitch peaks. *Journal of Phonetics* 22: 269–281.
- Hayashi, Makoto. 2010. An overview of the question–response system in Japanese. *Journal of Pragmatics* 42: 2685–2702.

- Hymes, Dell. 1974. *Foundations in Sociolinguistics*. Philadelphia, PA: University of Pennsylvania Press.
- Kearsley, Greg. 1976. Questions and question-asking in verbal discourse: A cross-disciplinary review. *Journal of Psycholinguistic Research* 5/4 : 355–375.
- Kiefer, Ferenc. 2000. *Jelentélmélet*. [Semantic Theory] Budapest: Corvina.
- Kiefer, Ferenc (ed.). 1983. *Questions and Answers*. Dordrecht & Boston: Reidel.
- Kugler, Nóra. 2000. A mondattan általános kérdései. [General issues of syntax]. In: Borbála Keszler (ed.), *Magyar grammatika*. [Hungarian Grammar]. Budapest: Nemzeti Tankönyvkiadó. 369–393.
- Markó, Alexandra. 2007. Kérdő funkciójú hanglejtésformák a spontán beszédben. [Intonation contours with interrogative function in spontaneous speech]. *Beszédkutatás 2007*: 59–74.
- Olaszy, Gábor. 2002. A magyar kérdés dallamformáinak és intenzitás szerkezetének fonetikai vizsgálata. [A phonetic analysis of the intonation and intensity structure of Hungarian questions]. *Beszédkutatás 2002*: 83–99.
- Rossano, Federico. 2010. Questioning and responding in Italian. *Journal of Pragmatics* 42 : 2756–2771.
- Schegloff, Emmanuel and Harvey Sacks. 1973. Opening up closings. *Semiotica* 7/4 : 289–327.
- Schirm, Anita. 2007. A kérdések pragmatikája [Pragmatics of questions]. In: Váradi, Tamás (ed.), *Alknyelvdok I. Alkalmazott Nyelvészeti Doktorandusz Konferencia* [1st Congress of Doctoral Students in Applied Linguistics]. Budapest: MTA Nyelvtudományi Intézet. 161–171. Retrieved on 2013-02-19 from [www.nytud.hu/alknyelvdok07/proceedings07/Schirm.pdf](http://www.nytud.hu/alknyelvdok07/proceedings07/Schirm.pdf).
- Searle, John. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Stivers, Tanya. 2010. An overview of the question–response system in American English conversation. *Journal of Pragmatics* 42 : 2772–2781.
- Stivers, Tanya and N. J. Enfield. 2010. A coding scheme for question–response sequences in conversation. *Journal of Pragmatics* 42 : 2620–2626.
- Varga, László. 2002a. *Intonation and Stress. Evidence from Hungarian*. Houndmills, Basingstoke: Palgrave Macmillan.
- Varga, László. 2002b. The intonation of monosyllabic Hungarian yes-no questions. *Acta Linguistica Hungarica* 49 : 307–320.
- Vion, Monique and Annie Colas. 2006. Pitch cues for the recognition of Yes-No questions in French. *Journal of Psycholinguistic Research* 35 : 427–445.
- Yoon, Kyung-Eun. 2010. Questions and responses in Korean conversation. *Journal of Pragmatics* 42 : 2782–2798.

Alexandra Markó  
 marko.alexandra@btk.elte.hu  
 Eötvös Loránd University  
 Budapest