PROSODIC CUES OF DISTINGUISHING NEUTRAL AND NON-NEUTRAL YES/NO QUESTIONS IN HUNGARIAN: THE ACOUSTICS OF SURPRISE

Katalin Mády, Uwe D. Reichel, Anna Kohári, Cecília Sarolta Molnár, Ádám Szalontai

Hungarian Research Centre for Linguistics
mady.katalin|uwe.reichel|kohari.anna|molnar.cecilia|szalontai.adam@nytud.hu

ABSTRACT

Hungarian information-seeking yes/no questions are often realised with a single rising-falling f0 contour throughout the utterance. A common way to indicate surprise in Hungarian is to put an emphasis on more than one lexical unit, which creates multiple rise-fall patterns in questions. This paper compares f0 contours of accent groups (AGs) in non-neutral yes/no questions expressing surprise with neutral information-seeking questions, both consisting of a single intonation phrase including at least two AGs and thus multiple f0 contours. Acoustic analysis was based on f0 parameters of the AG that best represent the difference according to a clustering-based incremental feature selection procedure. Non-neutral AGs were characterised by lower f0, but larger f0 range than those appearing in neutral questions. This is in line with previous perception experiments showing that surprise is encoded in f0 parameters in Hungarian.

Keywords: prosody, yes/no questions, surprise, feature selection, Hungarian

1. INTRODUCTION

Speakers can express their attitude or pragmatic bias by several means. A well-known example is non-neutral word order in exclamations with *wh*-words that differ from the neutral word order for interrogatives, e.g. *How large the conference room is!* vs. *How large is the conference room?* Another efficient way is to utilise prosody divergent from the neutral or canonical pattern. Additionally, bias in attitude can be marked by lexical items, e.g. discourse particles.

An interesting case for speaker bias is when the speaker utters a yes/no question that is not an information-seeking interrogative in its function, rather, it expresses that the information the speaker receives goes against their previous expectations. Reasons can be surprise, disbelief, incredulity or disapproving. According to the formal description in English, these questions have rising intonation similar to neutral yes/no questions, but they keep the declarative word order. Due to the additional pragmatic information (speaker attitude or bias), they are categorised as rising declaratives (see [1] for an overview).

The way Hungarian marks yes/no questions is different from English due to the fact that this sentence type is only marked by prosody, not by word order, i.e. syntax. Alternatively, the morphological marker *-e* attached to the predicate can be used, but it is restricted to certain regional and stylistic varieties and specific contexts and will not be discussed further in this paper.

The intonational phonology of neutral yes/no questions is described among others by Ladd [2], referred to as Eastern European question intonation. A neutral yes/no question with broad focus, i.e. with no specific emphasis on any constituent, is typically realised by a rising-falling intonation starting with a low pitch accent on the initial syllable of the first lexical unit, followed by a rise until the penultimate syllable and a fall on the last syllable. In terms of intonational phonology [2, 3], the tonal pattern is modelled as given in (1). (In colloquial speech, it is common to use a definite article before proper names in most Hungarian varieties.)

(1)
Meghívták a Melindát a moziba?
L* H- L%
invited-they the Melinda+ACC the cinema-to
'Was Melinda invited to the cinema?'

The tonal pattern is not sensitive to the presence of word stress in the lexical units following the last (here: only) accented word. The f0 maximum is expected on the penultimate syllable if it is preceded by the low pitch-accented one. When the penultima is accented itself, it carries the low pitch accent, and the rise-fall sequence is realised on the last syllable. If the last syllable is pitch-accented, the rise-fall is truncated into a single rise.

In Hungarian, the rise-fall pattern characterising neutral yes/no questions is also used in biased

questions with functions similar to the English rising declarative pattern. The most influential paper on the intonation pattern for Hungarian incredulous questions is by Varga [3] who models multiple rise-fall patterns as independent intonation phrases (IPs) containing the pattern L*HL% in each contour [3]. This is illustrated by the following sentence.

(2)

Meghívták	a	Melindát	a	moziba?
L* H- L%	L* H-		L%	L* H- L%
invited-they	the	Melinda+ACC	the	cinema-to

'Melinda was invited to the cinema?'

While this approach clearly has its benefits, i.e. it can account for the complex boundary tone specific to the yes/no question pattern, it leaves other questions open. One is that the rise-fall contours are not independent in their f0 values: there is an overall declination resulting in descending f0 maxima on the penultimate syllable of each prosodic unit. Another counterargument against assuming independent IPs is that it is highly unnatural to insert a pause before the accented L* syllable, as was also pointed out by Varga. We assume instead that the multiple f0 contours can be regarded as lower-level prosodic phrases, i.e. accentual phrases (AP) or intermediate phrases (ip). APs are part of the Hungarian prosody, as was shown by [4], while the existence of ips has not been investigated so far. We will leave the structural description of the above pattern aside for the present investigation and use accent groups (AG) as the domain of prosodic investigation (see Section 2 below). A further argument against regarding AGs as being equivalent to independent IPs is that it is possible to produce the same sentence with two non-final rises and a final rise-fall rather than with three rise-falls. This is in fact the only option if the definite article *a* is dropped before the second and third pitch-accented lexical unit. Since the phonological status of the high tones is unclear, the - sign signalling a phrase accent in the intermediate phrase is omitted here, while the utterance is regarded as a single IP.

(3)

Meghívták	Melindát	moziba?
L* H	L* H	L* H L%

'Melinda was invited to (the) cinema?'

No matter which tonal model is preferred in Hungarian neutral and non-neutral yes/no questions, there is agreement that surprise/incredulity is expressed by the presence of multiple accents in utterances that could be produced with a single pitch accent on the first lexical unit in the neutral question like in (1). Marking surprise by an intonation pattern different from neutral information-seeking questions was also observed in French [5]: surprise questions end less often with a rising contour, but their mean fundamental frequency (f0) and pitch range does not differ from neutral questions. In Estonian, surprise is manifested in the following acoustic parameters [6]: surprise questions are produced with higher initial pitch and a larger pitch range, but with lower mean f0 over the entire utterance and with more creaky voice.

For Hungarian, empirical results on the intonation of yes/no questions expressing surprise are available via a perception experiment by [7, 8]. Five gradually manipulated f0 curves represented a single five-syllabic word. Participants in the first experiment were asked to decide whether the utterance conveyed a request for confirmation or surprise; and in the second one, whether it expressed a question or surprise. The perception of surprise was elicited most reliably if the first part of the rise-fall was elevated, i.e. the pitch accent was realised with higher f0 – this is in line with findings for Estonian surprise questions. Contours with a rise-fall were more likely to be identified as confirmation- or information-seeking questions.

These results give only indirect hints to the acoustic characteristics of yes/no questions expressing surprise as opposed to the more broadly defined category of neutral (information- or confirmation-seeking) questions. Thus, it remains to clarify the following issues:

- Is there any formal difference between the f0 contours in neutral and non-neutral questions? Do they differ in shape or in f0 range?
- Do neutral questions have a more pronounced rise-fall f0 pattern than those expressing surprise similarly to previous perceptual evidence?

2. DATA

The dataset is based on the Budapest Games Corpus [9] that was developed in a similar manner to the Columbia Games Corpus [10]. The task-oriented dialogues induced a large variety of sentence types with manifold pragmatic functions. A total of 525 yes/no questions were found. These were split into two main categories: neutral and non-neutral. The neutral category contained questions ($n = 461$) whose primary function was seeking information or confirmation. The non-neutral category contained questions ($n = 64$) that expressed surprise or disbelief when speakers were confronted with information contrary to their expectations.

Questions with multiple f0 contours for which the context did not provide evidence for expressing surprise (e.g. no previous contradicting expectation could be detected) were categorised as neutral. Interlabeller agreement between two expert labellers was 100% for the categories neutral vs. non-neutral on a subset of 21 questions. IPs were segmented into accent groups (AG) by labelling the accented syllable, the non-accented parts and potential pauses.

Given that neutral questions with a single accent are often realised with a rising-falling f0 contour spanning over a longer sequence like in Example (1), a direct comparison between non-neutral and neutral questions seems reasonable if only neutral questions with multiple AGs are considered. Thus, the final data set was limited to IPs consisting of at least two AGs, leaving us with 233 questions (neutral: 195, non-neutral: 38) containing 370 non-final and 228 final AGs altogether.

3. METHODS

3.1. Feature extraction

F0 was extracted by autocorrelation (Praat 6.0.37 [11], sample rate 100 Hz; allowed f0 range between 50 and 400 Hz; default settings). Voiceless parts and f0 outliers were bridged by linear interpolation. Outliers were defined separately for each file as deviating more than three times the standard deviation from the f0 mean. The contour was then smoothed by Savitzky-Golay filtering using third order polynomials in 5 sample windows and transformed to semitones relative to a speaker-dependent base value b. b was set to the f0 median below the 5th percentile of an utterance and served to normalise f0 with respect to its overall level.

For prosodic feature extraction we applied the CoPaSul software version 1.0.3 [12] that decomposes the intonation of an utterance into a global intonation phrase and a local accent group component. For this study, the global component consisted of an f0 base-, mid- and topline that were fitted to the $[0, 1]$ time-normalised f0 contour in the intonation phrases (IPs). This robust fitting procedure is described and discussed in [13]. The midline was then subtracted from the f0 contour in order to remove the global component, and 3rd order polynomials were fitted to the $[-1, 1]$ time-normalised residual contour within the AG segments. These polynomial coefficients represent the local contour shapes on the AG level. Additionally, we extracted f0 residual summary statistics like mean and standard deviation, as well as f0 register and Gestalt features for the AGs. Register features measure AG-related aspects of local f0 level and range from a linear base-, mid-, and topline fit. Range is represented by a linear fit to the pointwise distances between top- and baseline. Gestalt features compare these register features between AG and IP in order to quantify how much and in what way an AG "pops out" of the underlying IP (see [12] for further details). Our feature pool comprised 32 features in total.

3.2. Feature selection

Our goal was to determine those features that best represent the AG characteristics in neutral and non-neutral questions for IP-final and non-final AGs separately. The statistic testing of such a large amount of features, 32 in our case, for each of the two conditions *IP-final* and *non-final* considerably increases the risk of type 1 errors. Therefore, we decided for an alternative approach which is clustering-based incremental feature selection. We applied this approach separately for the IP-final and non-final AGs.

Starting with zero features we iteratively added the feature that maximally increases the agreement between the clustering of the AG feature vectors and their classes neutral vs non-neutral. We applied the k-means algorithm with two clusters, a *k-means++* initialisation, and a constant random seed. As a scoring function that measures the amount of agreement between the clustering and the AG types we used the adjusted mutual information score (AMI) [14] between cluster IDs and question type. It measures the amount of uncertainty reduction about one of these variables if the values of the other variable are known. The AMI further corrects for chance-level agreement. As stopping criterion we defined the absence of any feature that would further increase the AMI score. Both clustering and AMI calculation were done with the Python *scikit-learn* package version 0.24.2 [15].

4. RESULTS

Figure 1 shows the feature selection result for the non-final and final AGs within the IP. For non-final AGs the closest agreement between AG features and question types was achieved with the features *ml_c0, rng_m*, and *rng_d_fin*. For final AGs the features maximising the agreement were *c0, ml_m, m, c3* and *rng_c0*. These features are described in Table 1.

The violin plots in Figures 2 and 3 for non-final and final AGs show how neutral and non-neutral AGs differ with respect to the selected features.

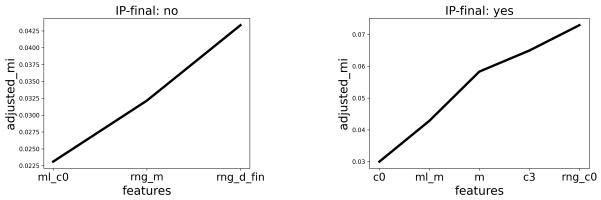

Figure 1: Feature selection results for non-final (left) and final AGs (right). See Table 1 for a description of these features.

Feature	Description
c0	offset coefficient of the local contour stylisation, that corresponds to the stylised f0 residual value at the midpoint of the AG
c3	hyperbolic polynomial coefficient of the local contour stylisation
m	arithmetic mean of the local f0 residual contour
ml_c0	offset coefficient (start level) of the linear AG midline fit
ml_m	arithmetic mean f0 of the AG midline
rng_c0	offset coefficient (start range) of the linear range fit to the pointwise distances between base- and topline in the AG
rng_d_fin	difference between AG and IP range at the end of the AG
rng_m	arithmetic mean of fitted range line in the AG

Table 1: Description of the AG intonation features resulting from incremental feature selection.

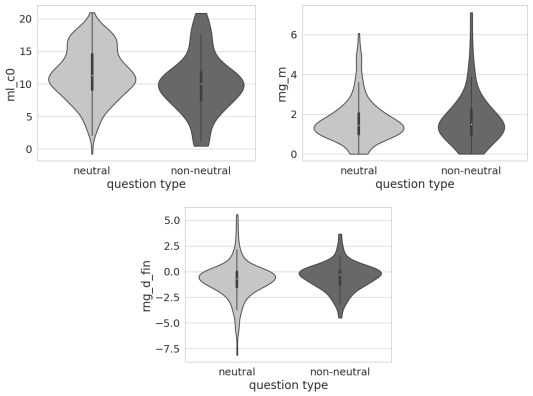

Figure 2: Distributions of selected features for **non-final** neutral and non-neutral AGs. See Table 1 for a description of these features.

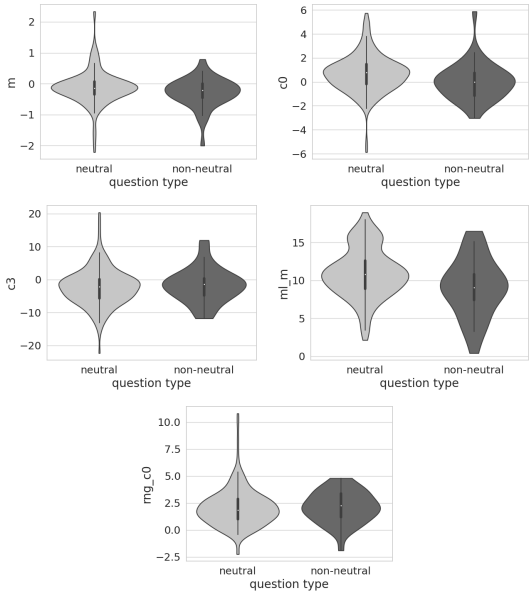

Figure 3: Distributions of selected features for **final** neutral and non-neutral AGs. See Table 1 for a description of these features.

5. DISCUSSION

Register can be expressed in terms of f0 *level* and *range* [16]. As Figures 2 and 3 show, this distinction is of relevance to describe both non-final and final, neutral and non-neutral AGs. Generally, non-neutral AGs can be characterised by a **lower f0 level** (in terms of f0 mean *m*, local contour offset coefficient *c0*, and midline mean *ml_m*), but **larger f0 range** (in terms of range offset coefficient *rng_c0*, and range mean *rng_m*). The difference might result from the fact that unlike in the perception experiment, our utterances contained at least two f0 contours. Furthermore, non-neutral AG intonation shows a **weaker downtrend tendency**, i.e. a smaller f0 slope which is expressed in terms of less negative hyperbolic coefficient values *c3*, which causes a less negative local f0 contour trend, and in terms of a larger range at the end of the AG *rng_d_fin*. These results only partly confirm the findings of [7, 8], according to whom questions perceived to indicate surprise have a larger f0 range and a *higher* f0 onset. At the same time, they are in line with findings for Estonian surprise questions.

According to the feature selection procedure, parameters representing the f0 contour shape such as *c2* were not chosen as contributing to the disctinction between the neutral and non-neutral category. This might be a result of the inhomogeneous contour shapes present in spontaneous speech.

6. ACKNOWLEDGEMENTS

This study was funded by the National Research, Development and Innovation Office, grants NKFIH K 135038 and PD 134775. We would also like to thank Beáta Gyuris and Hans-Martin Gärtner for their help to better understand the theoretical framework and the contextual interpretation of non-neutral questions.

7. REFERENCES

[1] B. Gyuris, "Thoughts on the semantics and pragmatics of rising declaratives in English and rise-fall declaratives in Hungarian," in *$K + K = 120$: Papers dedicated to László Kálmán and András Kornai on the occasion of their 60th birthdays*, B. Gyuris, K. Mády, and G. Recski, Eds. MTA Research Institute for Linguistics, 2019, pp. 247–280, http://www.nytud.hu/kk120/www_print/index.html.

[2] D. R. Ladd, *Intonational phonology, 2nd ed.* Cambridge: Cambridge University Press, 2008.

[3] L. Varga, "Boundary tones and the lack of intermediate phrase in Hungarian," *The Even Yearbook*, vol. 9, pp. 1–27, 2010.

[4] Š. Beňuš, U. D. Reichel, and K. Mády, "Modeling accentual phrase intonation in Slovak and Hungarian," in *Complex Visibles Out There. Proceedings of the Olomouc Linguistics Colloquium 2014.*, L. Veselovská and M. Janebová, Eds. Olomouc: Palacký University, 2014, pp. 677–689.

[5] A. Celle and M. Pélissier, "Surprise questions in spoken French," *Linguistics Vanguard: a Multimodal Journal for the Language Sciences*, vol. 8, no. 2, pp. 287–302, 2022. [Online]. Available: https://hal.science/hal-03133878

[6] H. Sahkai, E. L. Asu, and P. Lippus, "Prosodic characteristics of canonical and non-canonical questions in Estonian," in *Proc. Speech Prosody, Lisbon*, 2022, pp. 135–139.

[7] A. Kiss and A. Szalontai, "The form and meaning of hungarian confirmative and echo declarative questions," in *15. International Conference on the Structure of Hungarian*, Pécs, 2021, https://icsh15.netlify.app/.

[8] A. Kiss, "A magyar deklaratív kérdések karakterdallamának vizsgálata percepciós kísérlettel," in *Általános Nyelvészeti Tanulmányok*, K. Mády and A. Markó, Eds. Budapest: Akadémiai Kiadó, 2022, no. XXXIV, pp. 169–194.

[9] K. Mády, A. Kohári, U. D. Reichel, A. Szalontai, and P. Mihajlik, "The budapest games corpus," in *Beszédkutatás – Speech Research Conference*, T. E. Gráczi, V. Horváth, K. Juhász, A. Kohári, V. Krepsz, and K. Mády, Eds., Budapest, 2023, pp. 75–77.

[10] A. Gravano, v. Beňuš, H. Chávez, J. Hirschberg, and L. Wilcox, "On the role of context and prosody in the interpretation of 'okay'," in *Proc. 45th Annual Meeting of Association of Computational Linguistics*, Prague, 2007, pp. 800–807.

[11] P. Boersma and D. Weenink, "PRAAT, a system for doing phonetics by computer," Institute of Phonetic Sciences of the University of Amsterdam, Tech. Rep., 1999, 132–182.

[12] U. Reichel, *CoPaSul Manual – Contour-based parametric and superpositional intonation stylization*, RIL, MTA, Budapest, Hungary, 2016, https://arxiv.org/abs/1612.04765.

[13] U. Reichel and K. Mády, "Parameterization of F0 register and discontinuity to predict prosodic boundary strength in Hungarian spontaneous speech," in *Elektronische Sprachsignalverarbeitung 2013*, ser. Studientexte zur Sprachkommunikation, P. Wagner, Ed. Dresden, Germany: TUDpress, 2013, vol. 65, pp. 223–230.

[14] N. Xuan, V. Julien, S. Wales, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *Journal of Machine Learning Research*, vol. 11, pp. 2837–2854, 2010.

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[16] T. Rietveld and P. Vermillion, "Cues for perceived pitch register," *Phonetica*, vol. 60, pp. 261–272, 2003.