CONTRAST OR CONTEXT, THAT IS THE QUESTION

Francesco Cangemi[1], Martine Grice[2], Hae-Sung Jeon[3], Jane Setter[4]

[1,2]University of Cologne, [3]University of Central Lancashire, [4]University of Reading
fcangemi@uni-koeln.de, martine.grice@uni-koeln.de, HJeon1@uclan.ac.uk, j.e.setter@reading.ac.uk

ABSTRACT

In the past decades, most research on question intonation has relied on controlled experimental tasks for the production and perception of syntactically well-formed, pragmatically or syntactically ambiguous, unbiased and genuine requests for new information.

The exploration of a real-life interaction in East Midlands British English shows that such questions are neither frequent nor typical. Prosodic information does undoubtedly contribute to the identification of questions, as shown by comparing questionhood ratings, based on either audio recordings or orthographic transcriptions only. However, compared to the role of syntactic, lexical, contextual, sequential, and interactional information, the role of speech prosody is shown to be marginal.

These findings suggest that prosodic information is difficult to isolate or to manipulate independently. Therefore, question intonation can only be adequately studied by taking into account naturalistic, context-rich, spontaneous behaviour.

Keywords: Question, Intonation, Ecological validity

1. INTRODUCTION

Questions are ubiquitous in human life, from the private negotiations of early childhood to the world-changing endeavours of scientific research. Unsurprisingly, questions are also crucial in linguistics. They are at the crossroads of all areas of analysis (from phonetics to discourse), of all perspectives of inquiry (from typology to variationism), and of all contexts of application (from technology to education). As such, in the last century questions have attracted a large share of linguists' interest and have consequently been studied using a variety of methods. The evidence used for theorising about questions thus includes data from introspection (e.g. linguists' intuitions), from observation (e.g. records of naturally occurring interactions) and from experimentation (e.g. controlled elicitation in production and perception). Naturally, the weighting of these three types of evidence varies across areas, perspectives, and contexts of study.

In recent decades, linguists specifically interested in the phonetic and phonological aspects of questions seem to have favoured the use of evidence based on experimentation. This preference is not only implicitly attested in actual research practices (e.g. in Speech Prosody 2020, 90% of papers on questions relied on controlled elicitation in the laboratory context), but also explicitly voiced in publications on linguistic methodology. For instance, [1] emphasised how experiments allow researchers to control for confounding factors, isolate dimensions of interest, devise pertinent minimal pairs, efficiently elicit many such pairs, and thus explore both the formal and functional aspects of linguistic contrasts.

In this frame, the success of experimental methods is inherently linked to the notion of contrast, which has played a crucial role in modern linguistics since its earliest days [2:166]. However, equally early work suggests that linguistic patterns can only be understood when evaluated within their larger communicative context, including an exploration of everyday interactions [3:5]. This is particularly true for contrasts that involve meaning beyond the lexical level, where context plays a crucial role [4], as in the case of questions. Therefore, it is possible that by enforcing experimental control to maximise contrast, one might impoverish context to the point of altering the very nature of the observed phenomena. As a result, it has been suggested that question intonation and other prosodic phenomena should be studied via the observation of naturalistic, context-rich, spontaneous behaviour [5].

A synthetic approach is offered by [6] who support the integration of different types of evidence. This is akin to the program of cognitive ethology, according to which the "experimental simplification of a real-world situation is a reasonable research tactic when it *follows* careful real-world investigation" [7:337, original emphasis]. By giving temporal precedence to the observation of non-experimentally constrained interactions, the researcher can identify the most frequent and potentially most salient phenomena. These can then be used as cornerstone for further theoretical development to remain grounded in the interactionally relevant facts [8].

In this paper, we therefore set out to explore to what extent the questions typically employed in controlled experimental studies are representative of the questions we observe in real-world interactions. After providing a working definition of such typical questions (§2.1), we introduce the corpus (§2.2) and the labelling strategies (§2.3) used in this study. We then illustrate the distribution of different types of questions in our corpus (§3.1), the contribution of intonation to raters' confidence (§3.2) and a comparison of ratings based on different strains of information (§3.3). Then we interpret these findings as suggesting that, in real-world interactions, prosodic and non-prosodic information is intertwined in the production and perception of questions.

2. METHOD

2.1. Stipulations

Most studies agree with [9] in considering as typical a certain set of functional and formal features of questions used in production and perception experiments. Functional features include "the speaker emits a request for information" (unlike rhetorical questions) and "the speaker has no bias towards an expected answer" (unlike confirmation requests). Formal features include "the question can be uttered independently" (unlike tag questions) and "the question is syntactically well-formed" (unlike elliptical questions). Given the emphasis on contrastive analyses discussed in §1, these questions are usually compared to less typical items, including those which are not marked by particles or morphosyntactic devices (like declarative questions).

While more nuanced treatments of typicality in questionhood are available [10, 11], the characterisation above seems to fit at least the studies from the Speech Prosody 2020 conference mentioned above. Many such studies featured the production of read speech, with questions such as *Yǒurén chī níngméng me?* (Mandarin Chinese, "Does anyone eat lemons?") [12], *Kas keegi sööb sellerit?* (Estonian, "Does anyone eat celery?") [13], *Che cosa ti volevano servire?* (Salerno Italian, "What did they want to serve you?") [14, 15]. Question sentences were usually uttered after silently reading a contextualisation paragraph, which is assumed to prompt a variety of interpretations (e.g. unbiased new information seeking question or rhetorical question).

In the following, we therefore investigate the occurrence of syntactically well-formed, pragmatically or syntactically ambiguous, unbiased, genuine requests for new information.

2.2. Corpus

We analysed excerpts from an interaction lasting 2 hours and 15 minutes between two native speakers of an East Midlands variety of British English. The participants played an online adaptation of a miniature wargame [16] and discussed issues of game design after the match. *Zoom* [17, 18] was used to record separate audio tracks for the two participants, which in the following will be referred to as L(eft) and R(ight). *Open Broadcaster Software* [19] was used to record the gameplay video feed with the conversation as a single audio track. This video recording was subsequently edited and published by one of the participants [20].

The interaction was natural, in the sense of "'non-experimental', not co-produced with or provoked by the researcher" [21:530]. The first author provided technical assistance with the recordings, but the interaction would have taken place independently of this research. In fact, the two players L and R had independent and intrinsic motivations for the interaction. L and R were expected to play the match as part of an online tournament, and R intended to upload footage of the match to his *YouTube* channel. Since R is one of the co-developers of the wargame, he also intended to gather feedback from L concerning issues of game design and balance. After the recording, participants were asked for permission to use the recording for research purposes. They agreed to this and further noted that the interaction is publicly available on *YouTube*.

From the *Zoom* recording we extracted two excerpts of approximately 15 minutes. The first excerpt [20, 1:18:51-1:35:25] contains a game turn played by L and can be analysed as a structured goal-oriented interaction. Typically, L initiates game actions (e.g. declaring the target of an attack) and R reacts to these actions in a dynamic decision-making process. Both players also exchange other static information (e.g. properties of game pieces as written in the rulebook) and comment on unpredictable events (e.g. virtual dice rolls). The second excerpt [20, 2:04:28-2:19:37] contains a discussion between L and R on issues of game design and can be equated with a casual conversation between friends on a topic of mutual interest. The interaction is slightly asymmetrical, since R probes L's opinions in order to amend the game's rulebook based on this feedback. This difference in the interactional scope of the two excerpts was expected to maximise the range of types of observable questions. The corpus size was kept below 30 minutes to ensure the possibility of a thorough analysis.

2.3. Ratings

A research assistant segmented the audio recordings into interpausal units using *Praat* [22] and provided an orthographic transcription which was revised by the first author. Each excerpt was then segmented into 4 portions, which were each assigned to one of the remaining authors and to a fifth collaborator. All raters were phoneticians trained in prosodic analysis.

In the first annotation stage, the raters provided a coarse annotation by listening to the audio recording, reading the transcription, and flagging interpausal units which contained possible instances of questions. At this stage, we followed [10] and left annotation criteria intentionally vague. Raters were asked to monitor utterances which sounded like a question (i.e. prosodically), had the structure of a question (i.e. syntactically), or functioned like a question (i.e. interactionally).

In the second annotation stage, the interpausal units flagged as potentially containing a question were submitted again to the raters for a finer evaluation. For each item, we extracted the orthographic transcription and the audio signal of the interpausal unit of the pre-item context (up to 1.5 seconds before the item beginning) and of the post-item context (up to 1.5 seconds after the item end). This information was combined into three stimulus versions. Version (a) contained only the orthographic transcription of the item. Version (b) contained both the orthographic transcription and the audio recording of the item. Version (c) contained the orthographic transcription of the context and of the item, the audio recording of the pre-item and of the post-item context, but not of the item itself.

For each item, the three versions were presented for evaluation to three different raters, thus excluding the rater who flagged the item in the first annotation stage. In this way each rater read or listened to every item extracted from the corpus. In the second stage, annotation criteria were made explicit and approximated the definition provided in §2.1 above. Raters were asked to judge whether the item represented a “Core question”. This label was used to indicate an utterance in which the speaker “expresses an unbiased knowledge gap concerning new information, assumes that the listener is able and willing to fill this gap, directs the listener's attention to the gap, and invites the listener to fill it with a response”. Responses ranged from 1 (definitely not a Core question) to 9 (definitely a Core question), and were collected using a *Praat* script. The annotation files and the scripts for the extraction and presentation of stimuli are available in an online repository [23].

3. RESULTS

3.1. Typical questions are not typical

After transcription and segmentation, the two extracts contained 1147 interpausal units. In the first annotation stage, 140 of such units were rated as potentially containing a question. In the second rating stage, 122 of such units received a score of less than 8 (“the item is a Core question”) after being presented either with prosodic information (§2.3b) or with contextual information (§2.3c). This finding suggests that, at least from a distributional point of view, Core questions are not the most typical questions.

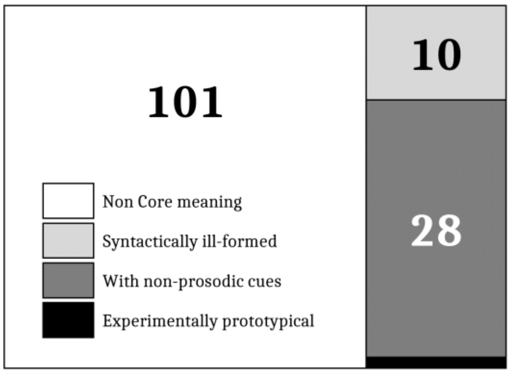

Figure 1: Core questionhood ratings.

We expanded the inclusion criteria to items with a rating of at least 7 (“probably a Core question”), thus filtering out only 101 cases (see Figure 1). Out of the remaining 39 items, 10 were not syntactically well-formed, and thus did not fit the typical features provided in §2.1. Of the remaining 29 items, 28 presented non-prosodic cues to questionhood, such as verb inversion or interrogative words. Only 1 item was rated as a Core question in the absence of non-prosodic marking. This finding raises the issue of how important intonation is in the identification of questionhood in natural interactions.

3.2. Prosodic information is important

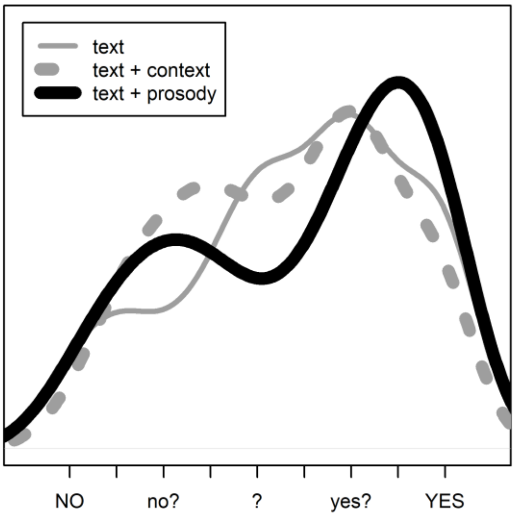

Figure 2: Questionhood ratings by type of stimulus.

We therefore explored how prosodic information contributes to the ratings, by visualising raters' responses separately for the three strands of information. In Figure 2, the thin grey line represents kernel density estimates [24] for responses to stimuli presented with orthographic transcription (§2.3a), the thick dashed grey line is used for stimuli with contextual information (§2.3c) and the thick black line for stimuli with prosodic information (§2.3b).

Compared to the grey density curves, the black curve for responses to stimuli with prosodic information shows greater separation between its two peaks. In other words, when raters were provided with prosodic information, they were more confident in their judgement of whether an item contains a Core question or not. Combined with the evidence from Figure 1, this finding suggests that prosodic information might not be essential to the identification of questions, but it might still play an important reinforcing role.

3.3. Non-prosodic information is crucial

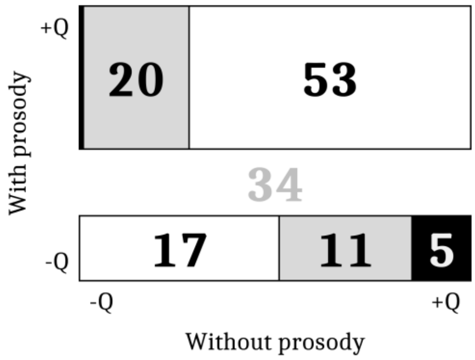

Figure 3: Agreement across questionhood ratings.

In order to further explore the contribution of prosodic information, we analysed the 74 items with positive Core question ratings (responses 7, 8 or 9, shortened to "+Q" below) separately from the 33 items with negative ratings (1, 2 or 3, "-Q"). Conservatively, in this analysis we ignore the 34 items which received intermediate ratings.

Figure 3 shows results along the y-axis for stimuli with prosodic information (§2.3b) and along the x-axis for the others. Out of 74 items rated as +Q based on prosodic information (top row), 53 were also rated as +Q based on textual or contextual information only. Only 1 was rated as -Q. Similarly, but less strikingly, of 33 items rated as -Q based on prosodic information (bottom row), 17 were rated as -Q based on textual and contextual information only, while 5 were rated as +Q. The fact that Core questions can be identified in most cases via textual and contextual cues demonstrates a high overlap between informational dimensions, and suggest that prosodic information might be difficult to isolate or to manipulate as a separate source of information.

4. DISCUSSION

Although the present study used a small set of data from one type of interaction, our findings highlight some potential drawbacks to the experimental study of question intonation. First, Figure 1 shows that the type of questions which are largely featured in experimental research might not necessarily correspond to the type of questions that are prevalent in everyday language use. This finding joins ends with research showing that syntactically well-formed, unbiased requests for new information might not be theoretically [11] or empirically [25] typical.

Second, and perhaps more radically, Figure 3 serves as a reminder that linguistic categories tend to be encoded redundantly [26], i.e. using different strands of information, which are then all used in perception by listeners. This casts a shadow on the assumption that different layers of information can be separated, isolated, manipulated and controlled in an experimental setup. This scepticism is compatible with views recognising the cohesive nature and function of prosody [27], against a mere superposition of segmental and suprasegmental levels [28]. It also echoes the body of work on prosodic constructions, where prosodic and lexical material are not seen as necessarily independent [29, 30].

While recognising the benefits of studying language in a controlled environment, it might thus be necessary to either verify the experimental findings using evidence from different sources [6, 31], or to ground experimentation on previous thorough observation of the phenomena of interest outside of the laboratory [7], where language is produced and perceived in a contextualised and meaningful way. Failure to do so runs the risk of divorcing research methods and goals from the reality in which they should be grounded [32], potentially leading to the atomistic, disconnected and often irreplicable body of findings [33, 34] that currently characterises research on question intonation.

ACKNOWLEDGEMENTS

FC conceptualised the study, collected and analysed the data, and wrote the first manuscript draft. MG, HJ and JS rated the stimuli and revised the manuscript. We are grateful to Hauke Lindstädt (segmentation, transcription) and Aviad Albert (additional ratings). Funding was obtained from the German Research Foundation (DFG), Project-ID 281511265, CRC 1252 "Prominence in Language".

REFERENCES

[1] Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics*, *38*(3), 329–336. https://doi.org/10.1016/j.wocn.2010.04.003

[2] de Saussure, F. (1916). *Cours de ling. générale*. Payot.

[3] Spitzer, L. (1921). *Italienische Kriegsgefangenenbriefe. Materialien zu einer Charakteristik der volkstümlichen italienischen Korrespondenz*. Peter Hanstein.

[4] Benveniste, É. (1974). *Problèmes de linguistique générale II*. Gallimard.

[5] Rischel, J. (1992). Formal linguistics and real speech. *Speech Communication*, *11*(4–5), 379–392. https://doi.org/10.1016/0167-6393(92)90043-7

[6] Wagner, P., Trouvain, J., Zimmerer, F. (2015). In defense of stylistic diversity in speech research. *Journal of Phonetics*, *48*, 1–12. https://doi.org/10.1016/j.wocn.2014.11.001

[7] Kingstone, A., Smilek, D., Eastwood, J. D. (2008). Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317–340. https://doi.org/10.1348/000712607X251243

[8] Albert, S., de Ruiter, J. P. (2018). Improving Human Interaction Research through Ecological Grounding. *Collabra: Psychology*, *4*(1), 24. https://doi.org/10.1525/collabra.132

[9] Haan, J. (2000). *Speaking of Questions. An Exploration of Dutch Question Intonation*. LOT.

[10] de Ruiter, J. P. (2012). *Questions: Formal, Functional and Interactional Perspectives*. Cambridge University Press. https://doi.org/10.1017/CBO9781139045414

[11] Ozerov, P. (2019). This is not an interrogative: The prosody of “wh-questions” in Hebrew and the sources of their questioning and rhetorical interpretations. *Language Sciences*, *72*, https://doi.org/10.1016/j.langsci.2018.12.004

[12] Zahner, K., Xu, M., Chen, Y., Dehé, N., Braun, B. (2020). The prosodic marking of rhetorical questions in Standard Chinese. *Speech Prosody 2020*, 389–393. https://doi.org/10.21437/SpeechProsody.2020-80

[13] Asu, E. L., Sahkai, H., Lippus, P. (2020). The prosody of rhetorical and information-seeking questions in Estonian: Preliminary results. *Speech Pros. 2020*, 381–384. https://doi.org/10.21437/SpeechProsody.2020-78

[14] Orrico, R., D’Imperio, M. (2020). Tonal specification of speaker commitment in Salerno Italian wh-questions. *Speech Prosody 2020*, 361–365. https://doi.org/10.21437/SpeechProsody.2020-74

[15] Prieto, P., Borràs-Comes, J., Roseano, P. (Coords.) (2010-2014). *Interactive Atlas of Romance Intonation*. http://prosodia.upf.edu/iari/

[16] The Ninth Age (2022). *Essence of War*. https://www.the-ninth-age.com/download/quick-play/

[17] Archibald, M. M., Ambagtsheer, R. C., Casey, M. G., Lawless, M. (2019). Using Zoom Videoconferencing for Qualitative Data Collection. *International Journal of Qualitative Methods*, *18*, 1609406919874596. https://doi.org/10.1177/1609406919874596

[18] Zoom Video Communications Inc. (2021). *Security guide*. https://explore.zoom.us/docs/doc/Zoom-Security-White-Paper.pdf

[19] OBS Studio Contributors (2021). *Open Broadcaster Software*. https://obsproject.com

[20] Agoners (2022). *The 9th Age EoW Advanced Battle: "This Is Why We BetaTest!!"*. https://youtu.be/VTkdztIq0OE

[21] ten Have, P. (2002). Ontology or methodology? Comments on Speer’s `natural’ and `contrived’ data: a sustainable distinction? *Discourse Studies*, *4*(4), 527–530. https://doi.org/10.1177/14614456020040040701

[22] Boersma, P. Weenink, D. (2022). *Praat: doing phonetics by computer*. https://www.praat.org/

[23] Cangemi, F. (2023). *23-08.* Supplementary material for “Contrast or context, that is the question”. https://osf.it/wp62b

[24] R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

[25] Enfield, N. J., Stivers, T., Levinson, S. C. (2010). Question–response sequences in conversation across ten languages. *Journal of Pragmatics*, *42*(10), 2615–2619. https://doi.org/10.1016/j.pragma.2010.04.001

[26] Winter, B. (2014). Spoken language achieves robustness and evolvability by exploiting degeneracy and neutrality. *BioEssays*, *36*(10), 960–967. https://doi.org/10.1002/bies.201400028

[27] Arvaniti, A. (2016). *Basic notions of prosody*. Talk at Aix Summer School on Prosody. https://aixprosody2016.weebly.com/program.html

[28] Roettger, T., Grice, M. (2019). The tune drives the text - Competing information channels of speech shape phonological systems. *Language Dynamics and Change*, *9*, 265–298. https://doi.org/10.1163/22105832-00902006

[29] Couper-Kuhlen, E., Selting, M. (1996). Towards an interactional perspective on prosody and a prosodic perspective on interaction. E. Couper-Kuhlen & M. Selting (eds.), *Prosody in Conversation: Interactional Studies* (pp 11-57). Cambridge University Press.

[30] Ward, N. G. (2019). *Prosodic Patterns in English Conversation*. Cambridge University Press. https://doi.org/10.1017/9781316848265

[31] Beckman, M. E. (1997). A Typology of Spontaneous Speech. Y. Sagisaka, N. Campbell, & N. Higuchi (eds.), *Computing Prosody: Computational Models for Processing Spontaneous Speech* (pp. 7–26). Springer US. https://doi.org/10.1007/978-1-4612-2258-3_2

[32] Neisser, U. (1978). Memory: What are the important questions? M. M. Gruneberg, P. E. Morris, & R. N. Sykes (eds.), *Practical aspects of memory* (pp. 3-24). Academic Press.

[33] Savino, E. (2012). The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association*, *42*(1), 23-48. https://doi.org/10.1017/S002510031100048X

[34] Garassino, D., D. Dipino, F. Cangemi (2022). Per un approccio multidimensionale allo studio dell’intonazione. In Schmid, S., Bernardasci, C., Dipino, D., Garassino, D., Negrinelli, S. & Pellegrino, E. (eds.), Speaker Individuality in Phonetics and Speech Sciences: Speech Technology and Forensic Applications. Milano: Officinaventuno. 219-242. https://doi.org/10.17469/O2108AISV000012