A case for systematic sound symbolism in pragmatics: The role of the first phoneme in question prediction in context

Anita Slonimska (anita.slonimska@istc.cnr.it)
CLS, Radboud University, Nijmegen, The Netherlands
Institute of Cognitive Sciences and Technologies, CNR, Rome, Italy

Seán G. Roberts (sean.roberts@bristol.ac.uk)
Department of Archaeology and Anthropology, University of Bristol, UK
Language and Cognition Department, Max Planck Institute for Psycholinguistics, The Netherlands

Abstract

Turn-taking in conversation is a cognitively demanding process that proceeds rapidly due to interlocutors utilizing a range of cues to aid prediction. In the present study we set out to test recent claims that content question words (also called *wh-words*) sound similar within languages as an adaptation to help listeners predict that a question is about to be asked. We test whether upcoming questions can be predicted based on the first phoneme of a turn and the prior context. We analyze the Switchboard corpus of English by means of a decision tree to test whether /w/ and /h/ are good statistical cues of upcoming questions in conversation. Based on the results, we perform a controlled experiment to test whether people really use these cues to recognize questions. In both studies we show that both the initial phoneme and the sequential context help predict questions. This contributes converging evidence that elements of languages adapt to pragmatic pressures applied during conversation.

Keywords: questions; wh-words; question words; turn-taking; speech-act recognition; question prediction

Introduction

People spend an average of 2-3 hours every day in conversation, producing around 1200 turns (Levinson, 2016). The structure of conversation, far from being chaotic, places specific constraints on speakers (Sacks, Schegloff & Jefferson, 1974). Recently, it has been recognized that these constraints have implications for processing and therefore for the way languages evolve (see Levinson, 2016). In this paper we explore a phenomenon at the interface of conversation, processing and cultural evolution.

Conversation progresses through exchanging bursts of information – mostly through use of language – that are orchestrated in consecutive turns produced by the speakers (Sacks et al., 1974). The surprising aspect of turn-taking is that it is orchestrated in a remarkably tight manner. Speakers strive to minimize gaps and overlaps between turns (Sacks et al., 1974), with the average gap length being only 200ms cross-culturally (Stivers et al., 2009; Kendrick & Torreira, 2015; Levinson & Torreira, 2015). Thus, while languages themselves differ, the pressure for rapid turn-taking is the same.

The surprising fact that turns are produced in such a tight window of time becomes even more puzzling if we take into account that it takes a minimum of 600ms to plan and begin uttering a single word (Schriefers, Meyer, Levelt, 1990; Levelt, 1993). In this context, one has to ask a question – how is it possible that the gap between turns is shorter than the planning of the response? The obvious answer is prediction (Sacks, Schegloff & Jefferson, 1974; Levinson, 2013). Listeners project what the current speaker will say and when their turn will end (Holler and Kendrick, 2015; Bögels & Torreira, 2015). Thus, the next speaker can start preparing their turn in advance so that it can be delivered on time.

Predicting the specific type of a speech act is extremely important as different speech acts have different social and cognitive pressures on speakers. For example, when we are greeted, the greeter expects a greeting in response. Or when we are asked a question, we are socially obliged to give an answer, and hesitations can lead to inferences about the intent of the responder (Kendrick & Torreira, 2015). Thus, social constraints put pressure on cognition to respond rapidly in interactive conversation. We suggest that languages should evolve to provide listeners with early cues that facilitate this process. Perhaps the context in which this would be most evident is in recognizing questions, to which we now turn.

Answering questions is a complex process involving understanding the question, retrieving or calculating the relevant answer and planning the response. Previous research suggests that the planning of the response starts as soon as an answer can be retrieved (Bögels, Magyari & Levinson, 2015; Bögels, Casillas, & Levinson, 2016; Barthel, Meyer & Levinson, 2017). However, even before planning their answers, speakers first have to recognize that they are being asked a question.

Gisladottir, Chwilla, & Levinson (2015) show that people can recognize the type of a speech act at an early stage if the preceding turns sufficiently constrain the context. For example, if I have just produced an initiating turn (like a greeting or asking a question), my interlocutor is most likely to produce a responding action (like an answer), rather than ask a question of their own. Therefore, one early cue as to whether a question will appear is the prior context.

Beyond that, there are also early cues in the question itself, before the turn can be identified as a question syntactically or semantically. Levinson (2013) suggests that question recognition is possible due to front-loading of the cues at the beginning of a turn. For example, questions can be recognized by early cues in intonation (Levinson, 2013), pitch (Sicoli et al. 2014) and eye-gaze (Rossano, Brown & Levinson, 2009; Rossano, 2012). Moreover, shifting question words to the initial position of the utterance (e.g., wh-movement in English) appears to be one of the most evident examples of front-loading (Levinson, 2013). Even when wh-movement is not permitted in the formal grammar of many languages, it is often evident in colloquial interactions (e.g. in Japanese, Levinson, 2013). Surprisingly, though, there is no quantitative research investigating whether this feature actually helps in question recognition.

Slonimska & Roberts (accepted) were the first to quantitatively assess whether question words, also called *wh-words*, are plausible candidates as a cue to content question recognition. They suggest that a systematic phonetic similarity between question words within a language could provide a cue for that. In other words, if question words tend to sound similar, it would be easier for the addressee to predict that a question is about to be asked, and they can prepare themselves accordingly. For example, in English many question words begin with /w/ (*what, why, where, when*), and in Latvian many begin with /k/ (*kas, kad, kur, kurš, kas, kāpēc*).

Even though there is some qualitative research arguing that there is no systematicity in question words (Cysouw, 2004), Slonimska & Roberts (accepted) show that there is a statistical tendency for question words to sound similar within languages. When they analyzed 266 languages the authors found that there is a higher similarity between the first phoneme of question words (within languages) than would be expected by chance, than other sets of words and also when controlling for historical factors. Accordingly, Slonimska & Roberts argue that this phenomenon constitutes a product of cultural evolution that is selected for due to its benefit in interaction – i.e., rapid question recognition. Their study, however, is based purely on observational data of word forms. This leaves several issues to be addressed before their claim can be supported. First, are phonological regularities in question words actually statistically good predictors of questions in conversation? Secondly, do people actually use these cues to recognize questions? Finally, what is the relationship between the use of these cues and the prior conversational context?

We address these issues by means of two studies. First, we explore a large corpus of natural conversations and subsequently use the insights from the corpus study to design an experiment in which we test the hypotheses in a controlled setting by using stimuli from the same corpus.

As such, the present project not only informs the theoretical field in regard to question recognition, but it also makes a case for a new approach to research – namely, by creating a synergy between ecologically valid corpus analysis and experimentally controlled quantitative insights into the phenomenon.

Corpus study

Method

To assess whether we can gain support for our hypotheses, we first carried out an exploratory corpus analysis of naturalistic data – i.e., spoken conversations. We addressed this by means of the method of binary decision trees, also known as recursive partitioning (Strobl, Malley, and Tutz, 2009). A binary decision tree represents the optimal series of yes-no questions that a rational agent would ask about predictor variables in order to estimate an outcome variable (see Roberts et al., 2015).

In the current study we are interested in whether the first phoneme of the turn (first predictor) and context of the previous turn (second predictor) would help in recognizing an incoming turn as a content question (outcome variable). Namely, we predicted that the data would be clustered in such way that specific first phoneme (/w/, /h/ versus other phonemes in English) of the current turn and specific type of previous turn (non-initiating turn versus initiating turn) would help identify whether the current turn was a question. Unlike regression frameworks, the predictor variables that a binary decision tree uses are not set by the researchers, but chosen by an algorithm in order to maximize performance and parsimony. It could pick any combination of phonemes as identifying factors if suggested by the data. Therefore, our prediction of the form of the tree is a strong one.

Materials and design. We used the Switchboard corpus (Calhoun et al., 2010) that consists of telephone conversations in American English. This corpus is transcribed and annotated in detail, including a division of utterances into sequential turns by Roberts et al. (2015). The data was prepared for the analysis in R and later analyzed by means of the package "party" (Hothorn, Hornik & Zeileis, 2006).

Each observation consisted of a transition between two turns between speaker A and speaker B. We specified the outcome variable – question – according to whether B's turn (i.e., current turn) was a question (content/open question) or not, according to the dialogue act annotation. We used the last speech act of A's turn (i.e., previous turn) for the second predictor variable specifying whether this turn was initiating or non-initiating (see Roberts et al., 2015). For example, B's turn was "What kind do you like to watch" - this was a turn that was a question and that started with /w/. The turn that preceded this question (i.e., A's turn) was "and uh you know there so there only a few that i that i like to watch routinely" – this was a statement (i.e., non-initiating turn).

We excluded the following fillers from the B's turn: *ahm, er, ah, hmm, oh, uh, aa, um, ow*. Then, the first phoneme from B's turns was extracted to create the predictor variable phoneme. This variable consisted of 34 unique phonemes (coded according to the transcription convention of Switchboard). Finally, we excluded all turns for which B's turn was a backchannel, considering that backchannel serves a monitoring rather than an informing function.

The final data included 9185 turns in total out of which 226 turns were content or open questions. Out of all turns, 1456 were initiating and 7729 were non-initiating turns. 1562 current urns (17%) started with /w/ or /h/.

For the analysis we had 2 predictor variables: context from the A's turn (initiating or non-initiating) and first phoneme of the B's turn (34 unique phonemes). The outcome variable was whether the current turn (i.e., B's turn) was a content/open question.

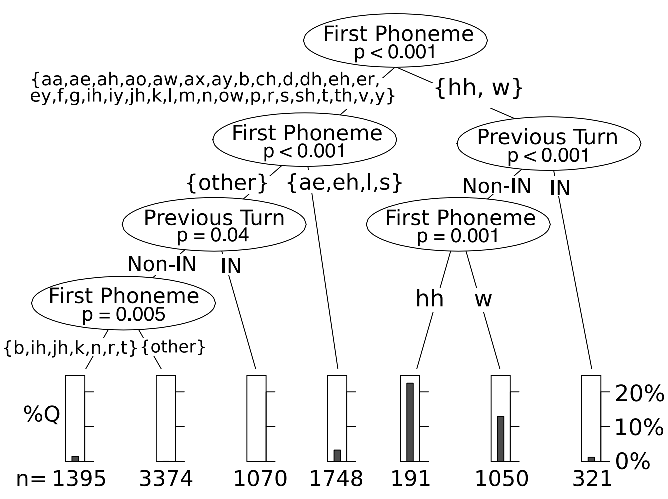
Results

The decision tree divides the data at each node of the tree starting from the top of the figure. Leaves of the tree at the bottom of the figure show the proportion of turns that are questions (see Fig.1).

The decision tree splits the data first based on the first phoneme of the turn. The exact division of the phonemes is as follows: /w/ and /h/ versus all the other phonemes, with the proportion of questions being higher for turns starting with /w/ and /h/. Thus, the decision tree, which is blind to our predictions, splits the data exactly in line with our predictions.

Following the branch that clusters the data on the right (/w/, /h/), the data is further clustered according to the type of the previous turn. If the previous turn was an initiating turn the proportion of question turns is considerably lower than if previous turn was not an initiating turn. If the

Figure 1: The decision tree of question turns split according to the sequential type of the previous turn and the first phoneme of the current turn. *Non-IN*: non-initiating turn, *IN*: initiating turn, phoneme transcription conventions come from the Switchboard corpus.



previous turn is not initiating, the data is further split into whether the phoneme of the current turn is /h/ or /w/. Note that proportion of questions is higher in /h/ (22%) leaf than in /w/ (13%). This may be because "well …", is often used as a filler at the beginning of a turn and thus decreases the overall proportion of questions in /w/ leaf. Moreover, there are more turns overall that start with /w/ than with /h/, therefore the proportion in /w/ leaf is also lower.

In regard to the data clusters on the left (turns starting with phonemes other than /w/ and /h/), it is evident that the proportion of question turns is extremely low in all leaves of the tree.

Overall, the analysis confirmed our initial hypotheses. Furthermore, based on the analysis we can also expect that the probability of a turn being a question will be additionally boosted if both cues are present – namely, if an incoming turn starts with /w/ or /h/ and the previous turn is non-initiating.

Experimental study

The corpus study suggested that the prior context and the initial phoneme of a turn helps identify questions statistically. The experimental study tests whether real people actually make use of these cues.

Method

Participants. For the experiment 25 participants (14 male, 11 female) were recruited. Participants' age ranged from 21 – 70 years (M = 32, SD = 11). All participants were native speakers of English but had various nationalities (e.g., American, British, Canadian, Australian, Indian, Latvian).

Materials and design. In this experiment participants listened to series of audio samples extracted from the Switchboard corpus. Each sample consisted of a context turn (initiating or non-initiating) produced by the first speaker and a response produced by the second speaker.

The *context turn* type could be either *initiating* (yes/no questions and wh-questions) or *non-initiating* (statements). The *response turn* type could be either *content questions* or *non-questions*. Each response turn was clipped to contain only the first phoneme, which could either be a *wh* phoneme (/w/ or /h/) or another phoneme. We therefore had the following fully crossed 2 x 2 x 2 design: *context type* (initiating/non-initiating) x *response type* (content question/other) x *response phoneme* (wh/other). In addition, the response turn could be blank (no audio, with context being initiating or non-initiating). This resulted in 10 conditions.

Table 1: Example of a 10 conditions consisting of 2 types of context turn (initiating/non - initiating) and 5 types of response turn.

Context turn		Response turn				
		/w/		Other		**Blank** no 2nd turn
		/w/ ques.	/w/ not quest.	not /w/ quest.	not /w/ non-quest.	
Not initial	***I do enjoy playing***	***Wh***[at is your…]	***W***[ell I wish…]	***D***[o you have…]	***Q***[uite a while…]	-
Initial	***And how did it go***	***Wh***[at is your…]	***W***[ell I wish…]	***D***[o you have…]	***Q***[uite a while…]	-

Context type was manipulated to test the effect of context and response phoneme was manipulated to test the effect of the first phoneme. Response type was manipulated so that we could assess whether the other question cues (e.g., raised pitch at the beginning of the question word) contribute in question prediction. The blank turn was added to establish a baseline for predicting an upcoming question without an initial phoneme.

We used the software Praat (Boersma & Weenink, 2014) to cut and concatenate each first turn with each second turn (e.g., (first turn: statement) + (second turn: /w/ from wh-question)). Subsequently, each turn pair was processed in the software Audacity (Mazzoni & Dannenberg, 2000) by adjusting a gap between the turns, so that the gap between first and second turn was 250ms.

We created 25 samples for each of the 10 conditions, resulting in 250 unique audio samples. These were split into 5 groups of 50 samples so that each context sample or response sample only appeared once inside each group.

Procedure. The experiment was presented via the online software Qualtrics (Snow & Mann, 2010). In each trial, a participant clicked a button to listen to a sample through headphones. Then they were asked to determine whether the second person would ask a question or not by means of completing a sentence "The Second turn is ____" on the screen by pressing one of two buttons: "not a question" or "a question". The experiment began with 2 practice trials ensuring that participants understood the task. Participants were assigned to an audio sample group and heard the samples from that group in a random order.

Results

We excluded 1 participant from the analysis due to the fact that this participant took 3 times longer than other participants to complete the experiment (38 minutes compared to an average of 12 minutes).

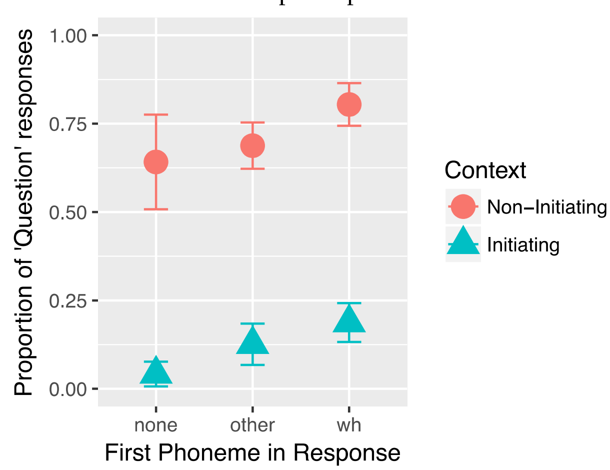
A logistic mixed model was used to predict whether the participant thought the response turn was a question (binary decision, yes or no, using the R package *lme4*, Bates et al, 2015). The predictor variables were *context* (initiating/non-initiating) and *phoneme* (wh, other, none). These predictors were coded as fixed effects and compared to a baseline model which included fixed effect of trial, random effect of context sample and phoneme sample, random effect of participant and random slopes for context and phoneme by participant.

There was a significant main effect of context ($\chi^2(1)$ = 45.74, p < .001). Participants were more likely to rate the turn as a question when preceded by a non-initiating context than an initiating context (see Fig.2).

There was a significant main effect of phoneme ($\chi^2(2)$ = 13.83, p < .001). Turns that started with *wh* phonemes were more likely to be rated as questions in comparison to turns starting with other phonemes or without the response from the second speaker. The model estimated that the probability of considering a turn a question was 90% for *wh* phonemes compared to 71% for other and 70% for none in non-initiating context. In initiating context this was 9% compared to 4% for other and 2% for none. There was no significant difference in question prediction between other phoneme and no response. Considering that there was only one variant of /h/ responses present in our stimuli, we ran analysis with these trials removed. There was no difference in the results with or without these trials.

Importantly, we also assessed whether participants could differentiate between the type of the response sample (a question or not) from which the phoneme was extracted. We found no effect of the response type ($\chi^2(1)$ = 0.11, p = .75).

Figure 2: Raw proportions of participants answering that an incoming turn is a question based on the previous context and the first phoneme of the incoming turn. Error bars indicate 95% CI of observations grouped within participants.



Thus, participants answered comparably to the phoneme samples that actually were questions and samples that were not questions. Most importantly, there was no interaction between response phoneme and the type of the response (χ^2= 0.008, p = 0.93). Thus, participants treated *wh* phonemes from real questions comparably to *wh* phonemes from other speech acts. These results suggest that participants are responding to the phoneme, not any other acoustic cue in the sample.

There was no significant interaction between context and phoneme ($\chi^2(2) = 1.34$, $p = .51$), although the trend was in the predicted direction.

Discussion and conclusions

In the present paper we set out to explore whether the first phoneme of a turn and the prior context can serve as a cue to question recognition. We found that both of these features contribute to this process. Although an effect of context was clearly expected, it was less certain whether there would be an effect of the first phoneme. This is the first experimental study supporting the claim of Slonimska & Roberts (accepted) that the first phoneme of question words can be used to predict an upcoming question.

We approached this topic from two different but mutually enhancing perspectives. We first assessed the hypothesis by analyzing natural conversations. Thus, we could look for patterns in ecologically valid data. The fact that the decision tree generated the same predictions as our hypothesis served as a sound basis for an experimental testing. Indeed, the samples from the corpus were used as experimental materials and the design was partly informed by the interaction between the two factors that the corpus study suggested. The hypotheses were also confirmed in the experiment, but there were two minor differences. First, the initial phoneme had a stronger effect than context in corpus study and vice versa in the experiment. Secondly, the corpus study predicted an interaction between initial phoneme and context, which was not found in the experiment. This may be because the probability of occurrence of various combinations is different in the corpus compared to the experiment, the experiment did not have enough statistical power, or more generally there is a difference between cues that are present in the data and ones that are actually used by people.

Another obvious difference between the two studies is that the speakers in the corpus had more prior context information than participants in the experimental study. Future experimental studies could include more extensive contextual information for the participants to be able to make predictions about the incoming turn.

Furthermore, the experimental participants were only passive listeners of the audio samples and their responses were not on-line. Future studies could take advantage of new paradigms to make it possible to combine interactive conversation with the use of controlled audio samples (e.g. Bögels, Magyari & Levinson, 2015).

Slonimska & Roberts (accepted) argue that question words tend to sound similar at the beginning of the word within a language to trigger question recognition. This leads to a prediction that /w/ should be a better cue than /h/, considering that there are more question words starting with /w/ than /h/. We found support for this in the corpus study. However, there was only one instance of /h/ phoneme in the experimental samples. We ran analyses with /h/ samples excluded and found no difference in the results. Therefore, although /w/ appears to boost question recognition, generalization to *wh* phonemes in English may not be warranted. Future studies could consider the differences between hearing /w/ and /h/ at the beginning of a turn in regard to question recognition.

It could be argued that the effect sizes in either study are too small to cause an evolutionary change in the language. However, we point out that even a small pressure would exert itself many times even in one conversation, and across cultural evolutionary time, small changes can accumulate to cause substantial changes.

Importantly, we advocate the virtuous cycle of looking for the phenomena in natural data, testing it in a controlled way and referring back to the real world. It can raise new questions and, most importantly, research can proceed in a more valid way than by using a single approach. This is clearly evident in our study - two approaches used in our study revealed differences that are important to account for, and which a single approach would have missed.

The findings in this paper are limited to English language and future research should continue exploring this cue in other languages, as well as diachronically. Only in this way can we be certain that this is not a single-language phenomenon or based on some idiosyncrasy of English but is actually a universal tendency. However, the puzzle remains - why else would question words sound so similar within so many languages (given that Slonimska & Roberts account for historical factors in their study and still find significant similarities)?

To summarize, by using different approaches in exploring the same topic we now have converging evidence for the question word similarity hypothesis: first, question words tend to sound similar within languages (Slonimska & Roberts, accepted); also, this phonetic cue can help in predicting questions in real conversations as shown in the corpus analysis; and finally people actually use this cue to predict questions when presented in a semi-natural setting. Thus, we suggest that the tendency for question words to sound similar is not a random occurrence, but might have evolved under a selective pressure to act as one of the early cues for question recognition in interactive conversation.

Acknowledgements

This research was supported by an European Research Council Advanced Grant No. 269484 INTERACT to Stephen Levinson and a Leverhulme early career fellowship to Seán Roberts (ECF-2016-435). We thank the Max Planck Institute for additional support.

References

Barthel, M., Meyer, A. S., & Levinson, S. C. (2017). Next Speakers Plan Their Turn Early and Speak after Turn-Final "Go-Signals". *Frontiers in Psychology*, 8, 393.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48

Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer. 7 http://www. fon. hum. uva. nl/praat/. Zugegriffen: 17.

Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*,52, 46-57.

Bögels, S., Casillas, M., & Levinson, S. C. (2016). To plan or to listen? The trade-off between comprehension and production in conversation. In the *Eighth Annual Meeting of the Society for the Neurobiology of Language*.

Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). Never Say No... How the Brain Interprets the Pregnant Pause in Conversation. *PloS one*, 10(12), e0145474.

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific reports*, 5, 12881.

Cysouw, M. (2004, February). Interrogative words: an exercise in lexical typology. In *Presentation presented at the Bantu grammar: description and theory workshop*, February (Vol. 13).

Enfield, N. J., Stivers, T., & Levinson, S. C. (2010). Question–response sequences in conversation across ten languages: An introduction. *Journal of Pragmatics*, 42(10), 2615-2619.

Gisladottir, R. S., Chwilla, D. J., & Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PloS one*, 10(3), e0120068.

Calhoun, S., Carletta, J., Brenier, J. M., Mayo, N., Jurafsky, D., Steedman, M., & Beaver, D. (2010). The NXT-format Switchboard Corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language Resources and Evaluation*, 44(4), 387-419.

Holler, J., & Kendrick, K. H. (2015). Unaddressed participants' gaze in multi-person interaction: optimizing recipiency. *Front. Psychol*, 6(98), 10-3389.

Holler, J., Kendrick, K. H., Casillas, M., & Levinson, S. C. (2016). Turn-Taking in Human Communicative Interaction. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-825-2

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3), 651-674.

Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: a quantitative study. *Discourse Processes*, 52(4), 255-289.

Levelt, W. J. (1993). *Speaking: From intention to articulation* (Vol. 1). MIT press.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6, 731.

Levinson, S. C. (2013). Action formation and ascription. In *The handbook of conversation analysis* (pp. 103-130). Wiley-Blackwell.

Levinson, S. C. (2016). Speech acts. In Y. Huang (Ed.), *Pragmatics*. Advance online publication. Oxford: Oxford University Press.

Mazzoni, D., & Dannenberg, R. (2000). Audacity (software). The Audacity Team, Pittsburg, PA, USA.

Roberts, S. G., Torreira, F., & Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: a corpus study. *Frontiers in Psychology*, 6, 509.

Rossano, F. (2012). Gaze in conversation. In J. Sidnell, & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 308-329). Malden, MA: Wiley-Blackwell.

Rossano, F., Brown, P., & Levinson, S. C. (2009). Gaze, questioning and culture. *Conversation analysis: Comparative perspectives*, 187-249.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 696-735.

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies.*Journal of Memory and Language*, 29(1), 86-102.

Sicoli, M. A., Stivers, T., Enfield, N. J., & Levinson, S. C. (2014). Marked initial pitch in questions signals marked communicative function. *Language and Speech*, 58(2), 204-223.

Snow, J., & Mann, M. (2013). *Qualtrics survey software: handbook for research professionals.*

Slonimska & Roberts (accepted). A case for systematic sound symbolism in pragmatics: universals in wh-words. *Journal of Pragmatics*.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587- 10592.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*,14(4), 323.