Interspeech 2018
2-6 September 2018, Hyderabad, India

A first investigation of the timing of turn-taking in Ruuli

Tuarik Buanzur[1], *Margaret Zellers*[1], *Saudah Namyalo*[2], *Alena Witzlack-Makarevich*[1]

[1]Institut für Skandinavistik, Frisistik, und Allgemeine Sprachwissenschaft,
University of Kiel, Germany
[2]Department of African Languages, Makerere University, Kampala, Uganda

stu102801@mail.uni-kiel.de, mzellers@isfas.uni-kiel.de, snamyalo@chuss.mak.ac.ug,
awitzlack@isfas.uni-kiel.de

Abstract

Turn-taking behavior in conversation is reported to be universal among cultures, although the language-specific means used to accomplish smooth turn-taking are likely to differ. Previous studies investigating turn-taking have primarily focused on languages which are already heavily-studied. The current work investigates the timing of turn-taking in question-response sequences in naturalistic conversations in Ruuli, an under-studied Bantu language spoken in Uganda. We extracted sequences involving wh-questions and polar questions, and measured the duration of the gap or overlap between questions and their following responses, additionally differentiating between different response types such as affirmative (i.e. type-conforming) or negative (i.e. non-type-conforming) responses to polar questions. We find that the timing of responses to various question types in Ruuli is consistent with timings that have been reported for a variety of other languages, with a mean gap duration between questions and responses of around 259 ms. Our findings thus emphasize the universal nature of turn-taking behavior in human interaction, despite Ruuli's substantial structural differences from languages in which turn-taking has been previously studied.

Index Terms: turn-taking, spontaneous speech, questions, Bantu, Ruuli

1. Introduction

Speakers in conversations mostly organize their speech in a way that conforms to a set of turn-taking rules which prioritize having one person talk at a time, and generally minimize both overlapping speech as well as long silent pauses or "gaps". Work by [1] and [2] has shown that the most frequent duration of silent between-turn gaps in a variety of languages is about 200 milliseconds (ms). Overlapping speech is also minimized, taking up as little as 5-10% of speaking time [3, 4, 5]. Additionally, different types of questions can lead to different answer timings, with, for example, polar questions being answered more quickly than wh-questions in general [6]. Furthermore, the appropriate timing of turn-taking is not simply a matter of style or preference; rather, silent pauses and overlaps can communicate additional information. "Dispreferred" responses, such as disagreements with an assessment [7], or non-type-conforming responses to questions (i.e. responses that do not have the same polarity as the question, cf. [8]), have been reported to be produced after a substantial delay on the part of the speaker [3]. Some recent large-scale quantitative studies have confirmed these reports. [1] find a significant difference in gap duration preceding confirmations (preferred) compared to disconfirmations (dispreferred), with significantly longer gaps preceding disconfirmations in 7 of the 10 languages they investigate. [9] report that the probability of a dispreferred conversational action increases with increasing gap duration, with actions following gaps of 700 ms or more particularly likely to be dispreferred. Similarly, [10] find that perceived willingness of a speaker to comply with a request (a preferred action) decreases after a gap of 600 ms or more.

The set of languages in which turn-taking has been studied is thus far fairly limited and has been heavily concentrated on languages of the Indo-European family. A few studies have investigated in more depth such geographically, genealogically and typologically different languages as Garrwa [11, 12], Baka and Bakwele [13], and Caribbean Creole [14], finding broad similarities in turn-taking behavior between these languages and other more heavily-studied languages, while suggesting that timing requirements may in some cases be more relaxed; that is, longer gaps may not lead to the interpretation of a following dispreferred action. [12] argues, however, that such a relaxation of timing requirements can also be found in appropriate social circumstances in e.g. Australian English conversations. It is clear that there is a substantial gap in current knowledge regarding the degree to which speakers of different languages display different turn-taking behaviors, such as variation in timing between conversational turns. The current paper thus investigates turn-taking, with a specific focus on question-response sequences, in an understudied Bantu language, Ruuli.

Ruuli (or Luruuli/Lunyara, ISO 639-3: ruc) is a Great Lakes Bantu language spoken mainly in the Nakasongola and Kayunga districts of central Uganda. The number of ethnic Baruuli/Banyara is 190,000 according to the 2014 census. The number of actual speakers is difficult to estimate at the present moment. [15] classify the language as threatened, i.e. the language is used for face-to-face communication within all generations, but it is losing users. The data used in the present paper were collected within the project *A comprehensive bilingual talking Luruuli/Lunyara-English dictionary with a descriptive basic grammar for language revitalisation and enhancement of mothertongue based education* (Volkswagen Foundation). Before the beginning of this project no description of this language existed.

Ruuli is a typical Bantu language and is thus typologically rather distinct from the more familiar Indo-European languages. The dominant constituent order is SVO, though other constituent orders are possible, as in (2) and (3). Nominal and verbal inflectional morphology is primarily prefixing. Nominal morphology is characterized by an intricate system of noun class prefixes. Every noun belongs to one of twenty nominal classes, which determine both the shape of nominal prefixes on nouns, as well as the shape of the agreement prefixes on dependents in a noun phrase, as e.g. on the numeral modifier 'four' in (1), on the verb, as well as on a number of other constituents,

10.21437/Interspeech.2018-1254

such as the question word 'how' in (4), which agrees with the implied subject *omusyetete* weaver ' bird' of class 3 (cf. [16]). The verb in Ruuli has about nine prefix slots and five suffix slots (the final analysis of the verb morphology is still pending). The verb of an independent clause obligatorily shows agreement with its subject, as well as optionally with its objects.

(1) *n-a-lek-ere=yo abaana ba-nai.*
1sgS-PST-leave-PFV-LOC children(2) 2-four
'I left four children there.'

Polar questions in Ruuli are not morphologically marked and the constituent order is identical to that of declarative sentences. A distinct intonation pattern, as well as the context are the sole indications of polar questions. An example of a polar question is given in (2), (3) provides a positive answer to it.

(2) *Omusyetete o-gu-maite?*
mousebird(3) 2sgS-3O-know.PFV
'Do you know the mousebird?'

(3) *Omusyetete n-gu-maite.*
mousebird(3) 1sgS-3O-know.PFV
'I know the mousebird.'

Wh-questions have a question word. Its position varies between the more common in situ position, as in (4), and the fronted position, as in (5). Individual question words have a preference for one of the two positions.

(4) *gu-naab gu-tyai*
3S-bathe 3-how
'How does it bathe?'

(5) *lwaki ba-ku-lu-yemba*
why 3plS-PROG-11O-sing
'Why do they sing it?'

2. Data and methodology

Following [1] and [6], we selected question-response sequences from the corpus of Ruuli data collected by the project described in Section 1.

2.1. Dataset

The corpus recordings were made using audio and video field recorders (Olympus LS-10, Zoom H4n Pro, Zoom Q8). Files are available in 16bit/44,100 Hertz stereo as .wav-files. Given that the primary aims of the recordings were morphosyntactic analysis and dictionary compilation, there is no separation of speakers by channels as is often the case for data collected under laboratory conditions. Instead, the recordings were mostly done outside in front of houses, accompanied by substantial village background noise from people, animals etc. Thus, our data represent Ruuli speakers in their familiar environment having naturalistic conversations.

Only a small portion of the corpus was used for the present study: we annotated 10 audio files with a length of almost 6 hours in total. The files consist of conversations containing mostly two and in a few cases three speakers. The recordings were made in six different locations. The total number of participants is 15, aged between 39 and 83 at the time of recording. Four of them are female, eleven are male. Although Ruuli speakers live in a multilingual environment—most Ruuli speakers also speak Ganda and/or English, as well as some other Bantu languages of Uganda—all participants acquired Ruuli as their first language. The topic of the conversation was not restricted, so the speakers talked about topics including culture, politics, and childhood memories, as well as local flora and fauna.

2.2. Selection and annotation

Orthographic transcriptions and translations of the recordings were made by language consultants. These original transcriptions and translations were then revised by the project members with the assistance of native speakers. The annotations for the current study were done using Praat [17]. To identify question-response sequences, a simple query of all question marks ("?") in the transcription was done in the first step. To accept the turns before and after the question marks as question-response sequence, two conditions had to be fulfilled: First, after "?" a speaker change must have taken place. Second, the turn of this speaker must have been in direct response to the turn before the "?" in such a way that the turn could be evaluated as a reaction to the first turn. These formal criteria allowed us to find question-response sequences in a language that is still in the beginning stages of documentation, and thus does not yet have a complete linguistically-annotated corpus, as is the case for Ruuli. By applying the formal criteria, we were able to limit our dataset to clear-cut cases of questions; turns with interrogative structure which were rhetoric in nature, or invited further clarifications or new topics, were excluded.

Questions were classified as wh-questions if there was an appropriate question word in Ruuli and English and otherwise as a polar question or an alternative question if different alternative answers were given in the formulation, following [1]. Polar questions which had the function of a repair initiation were labeled as "inquiry" and together with alternative questions they were excluded from the present study (due to a low number of relevant examples). For polar questions we also evaluated the response turns thus creating three different possible values for the response which are "affirmative" (i.e. type-conforming), "negative" (i.e. non-type-conforming), or "unclear" with respect to the question turn. Polar questions with an "unclear" response were not used in the analysis at hand. All annotations and evaluation decisions were made predominantly by the first author, and a second annotator verified more than one third of the labels. Labels and data were excluded from further analysis when the two annotators did not agree with each other about the labelling decision.

Transitions in question-response sequences were labelled as "gap", "overlap" or "0-gap". Durations were measured between the last element of the question turn, and the first element of the response turn, including particles, in breaths and clicks as part of the turn of the respective speaker. After labelling, time measurements were done automatically using a Praat script. Silent pauses between turns longer than 120 ms were labelled as "gap". Overlaps of question and response turns were labelled as "overlap" if the simultaneous sequences of the two turns lasted more than 120 ms. Simultaneous sequences and silent pauses that lasted less than 120 ms were analyzed as "0-gap", or no-gap-no-overlap, following [18]. With reference to the noisy recording environment mentioned under Section 2.1, a silent pause here means that neither of the two relevant speakers was audible to the annotator using headphones, and that there was no visual cue in the Praat window that showed evidence for a speech signal such as voice-dependent periodicity or frequency-specific friction noise which differed from the random noise in the background. It is possible that the aforementioned recording conditions caused some inconsistencies in the annotations; however, these are likely to be random rather than systematic, and were in any case unavoidable.

Table 1: *Distribution of gaps, no-gap-no-overlaps (0-gap), and overlaps, in polar questions with type-conforming (+) and non-type-conforming (-) responses, and wh-questions.*

	Gap	0-Gap	Overlap
Polar (+)	24	35	12
Polar (-)	22	9	3
Wh	75	55	26

Table 2: *Mean duration of gaps in polar questions with type-conforming (+) and non-type-conforming (-) responses, and wh-questions.*

	Gap duration (sec)
Polar (+)	0.240
Polar (-)	0.302
Wh	0.280

To sum up, the annotations done in Praat include orthographic transcriptions of question-response sequences, containing the respective turns in Ruuli and English, question type labels, transition labels, response labels for polar questions and a label to identify which speaker is talking.

The total number of questions originally identified in the current study is 315. After removing cases in which annotators were uncertain about the format of the question or response (e.g. whether a question was truly information-seeking vs. simply rhetorical, or what the polarity of a response was), the number of transitions used in the analysis is 261.

3. Analysis and results

The first portion of the analysis investigates the distribution of silent gaps, no-gap-no-overlaps, and overlaps occurring between questions and responses of three different types: polar questions with type-conforming responses, polar questions with non-type-conforming responses, and wh-questions. The second portion then investigates in more detail the duration of perceptible gaps and overlaps in these three question-response pair types.

3.1. Distribution of gaps and overlaps

The distribution of question-response sequences presenting with gaps, no-gap-no-overlaps, and overlaps, is shown in Table 1. A chi-square test shows that the distribution is significantly different from a random distribution (χ^2(4, N=262) = 10.053, p<.05). In particular, wh-questions and polar questions with a non-type-conforming response are relatively more likely to be followed by a gap, while polar questions with a type-conforming response are more likely to have their response be timed with no gap or overlap. Overlaps occur in 41 cases overall, or 15.7% of transitions; this is within the range of values reported by [3] (5% of cases) and [2] (20% of cases).

3.2. Duration of gaps and overlaps

The overall distribution of gap duration in the data (including gaps which were shorter than 120 ms and thus later classified as no-gap-no-overlap) is shown in Figure 1. The mean duration of all gaps in the data is 259 ms, while the density curve, indicating the most frequent values, peaks at around 125 ms (i.e. a just-noticeable gap, cf. [18]). These values are consistent with the values reported for a number of other languages by [1].

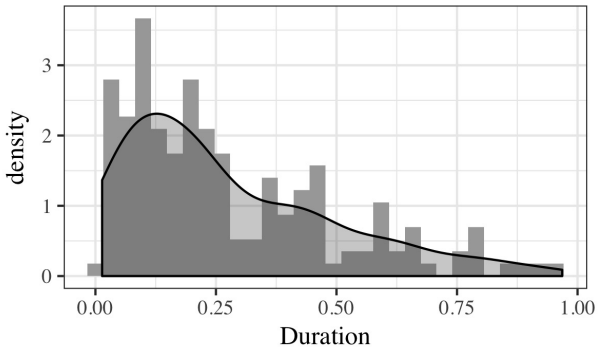

Figure 1: *Histogram of all measured gap durations (including short values which were later classified as no-gap-no-overlap).*

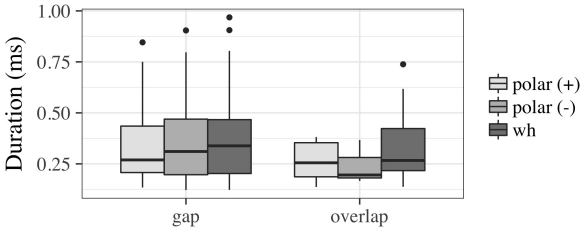

Figure 2: *Gap and overlap durations in polar questions with type-conforming (+) and non-type-conforming (-) responses, and wh-questions.*

Figure 2 shows the durations of gaps and overlaps in polar questions and wh-questions. As shown in Table 2, gaps following wh-questions and polar questions with non-type-conforming responses tend to be longer than gaps following polar questions with type-conforming responses. This trend is in the same direction as has been reported for other data by [1, 6], but a linear mixed model does not attain statistical significance in the current data, likely because of the relatively small size of the dataset.

4. Discussion

We investigate the timing of responses to polar and wh-questions in the under-studied Bantu language Ruuli. We find that, despite typological differences between Ruuli and other languages that have been studied with regard to turn-taking, the timing of responses to questions in Ruuli is comparable to what has been reported across a variety of languages. Additionally, we find that different question and response types are associated with different timings of responses in ways that are also consistent with previous reports in the literature. Our findings thus emphasize the universal nature of the turn-taking system for conversational interaction, despite cross-linguistic variability.

4.1. Question type and response timing

[6] investigated the timing of responses to different types of questions in English and Swedish. They found that wh-questions tended to be answered more slowly than polar ques-

tions overall, and that within polar questions, affirmative answers came more quickly than negative or dispreferred answers. These differences were found to be statistically significant for English, with a non-significant trend in the same direction in the Swedish data. They interpreted this difference as arising from the different size of the datasets, since the English set was almost ten times as large as the Swedish set. [1], with datasets closer to the size of the one used in this study, also only find a significant difference between confirming (preferred) and disconfirming (dispreferred) responses in 7 out of the 10 languages they investigate. Our dataset is even smaller than those used in these other studies, so it is perhaps unsurprising that we do not attain a statistically significant result for the linear mixed model. However, we find a trend for gap duration differences in the same direction as reported by [1] and [6]; that is, that type-conforming responses to polar questions come more quickly than non-type-conforming responses or responses to wh-questions. Furthermore, the significant chi-square test (Sec. 3.1) shows that type-conforming responses to polar questions are more likely to be produced with no gap or overlap, compared to non-type-conforming responses to polar questions or responses to wh-questions, which are more likely to arise with a gap.

Previous research has shown that longer gaps typically precede dispreferred responses such as disagreement, reluctance, or simply responses which do not conform to the polarity of a question [3, 7, 8]. Thus, longer gaps could be treated as a communicative strategy used intentionally by a speaker. Alternatively, it has been suggested that longer gaps preceding responses to wh-questions as compared to polar questions could result from a relatively higher cognitive load, due to the fact that the response might be more informationally dense, or cannot recycle as much material from the preceding question [6]. Additionally, some time is required for articulatory planning and launching of speech [5]. In Ruuli in particular, wh-words occurring in situ are mostly found sentence-finally; this may lead to a relatively late interpretation of the turn as a question, thus delaying responses to such wh-questions. We have not yet analyzed the questions included in the current study with regard to possible differences between wh-questions with fronted vs. in situ wh-words.

4.2. Linguistic differences and linguistic universals

Our finding that the average gap duration between questions and responses in Ruuli is approximately 259 ms, with a tendency for preferred responses to polar questions to come more quickly compared to dispreferred responses or responses to wh-questions, contributes additional evidence in support of the growing consensus that the turn-taking system, while involving some language-specific variation in exact details of timing, is fundamentally universal. As discussed by [1] and [5], while the language-specific preferences for appropriate timing between turns may differ, the factors that influence whether longer or shorter gaps are produced—such as preferred vs. dispreferred responses—have consistent effects across languages.

5. Conclusions and future directions

We find that, in common with many other typologically different languages, Ruuli tends to have a gap of on average 259 ms between questions and responses, with a tendency for type-conforming responses to polar questions to come more quickly than responses to other types of questions. Our expansion of the study of turn-taking in conversation to this Bantu language lends additional support to the idea that structured turn-taking is a linguistic universal. The continued investigation of less-studied languages such as Ruuli will shed more light on this matter.

The current study was intended as an initial investigation and has thus taken a fairly simple approach to the question of the timing of turn-taking; we limited our dataset to tokens which were unambiguous, and used only polar and wh-questions. Future research will investigate a wider range of turn pairs, taking into account aspects of linguistic structure including location of wh-words, as well as investigating the interplay of Ruuli's tone system with possible prosodic cues leading to the perception of turn ends.

6. Acknowledgements and abbreviations

The work of Margaret Zellers was supported by a start-up grant from the University of Kiel. The work of Saudah Namyalo and Alena Witzlack-Makarevich was supported by the project *A comprehensive bilingual talking Luruuli/Lunyara-English dictionary with a descriptive basic grammar for language revitalisation and enhancement of mothertongue based education*, funded by the Volkswagen Foundation.

Examples (1)–(5) use the following abbreviations: *1sg* 1st person singular, *2sg* 2nd person singular, *3pl* 3rd person plural, *2* noun class 2, *3* noun class 3, LOC locative, O direct object, PROG progressive, PST past, PFV perfective, S subject.

7. References

[1] T. Stivers, N. J. Enfield, P. Brown, C. Englert, M. Hayashi, T. Heinemann, G. Hoymann, F. Rossano, J. P. de Ruiter, K.-E. Yoon, and S. Levinson, "Universals and cultural variation in turn-taking in conversation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 26, pp. 10 587–10 592, 2009.

[2] M. Heldner and J. Edlund, "Pauses, gaps and overlaps in conversation," *Journal of Phonetics*, vol. 38, pp. 555–568, 2010.

[3] S. C. Levinson, *Pragmatics*. Cambridge, UK: Cambridge University Press, 1986.

[4] E. Shriberg, A. Stolcke, and D. Baron, "Observations on overlap: Findings and implications for automatic processing of multi-party conversation," in *Proceedings of Seventh European Conference on Speech Communication and Technology*, 2001.

[5] S. Levinson and F. Torreira, "Timing in turn-taking and its implications for processing models of language," *Frontiers in Psychology*, vol. 6, p. 731, 2015.

[6] S. Strömbergsson, A. Hjalmarsson, J. Edlund, and D. House, "Timing responses to questions in dialogue," in *Proceedings of 14th Interspeech, Lyon, France*, 2013, pp. 2584–2588.

[7] R. Ogden, "Phonetics and social action in agreements and disagreements," *Journal of Pragmatics*, vol. 38, pp. 1752–1775, 2006.

[8] G. Raymond, "Grammar and social organization: yes/no interrogatives and the structure of responding," *American Sociological Review*, vol. 68, no. 6, pp. 939–967, 2003.

[9] K. Kendrick and F. Torreira, "The timing and construction of preference: a quantitative study," *Discourse Processes*, vol. 52, pp. 255–289, 2015.

[10] L. S. Kohtz and O. Niebuhr, "How long is too long? How pause features after requests affect the perceived willingness of affirmative answers," in *Proceedings of 18th Interspeech, Stockholm, Sweden*, 2017, pp. 3792–3796.

[11] I. Mushin and R. Gardner, "Silence is talk: conversational silence in Australian Aboriginal talk-in-interaction," *Journal of Pragmatics*, vol. 41, no. 10, pp. 2033–2052, 2009.

[12] R. Gardner and I. Mushin, "Expanded transition spaces: the case of Garrwa," *Frontiers in Psychology*, vol. 6, no. 251, pp. 1–14, 2015.

[13] D. Kimura, "Utterance overlap and long silence among the Baka pygmies: comparison with Bantu farmers and Japanese university students," *African Study Monographs: Supplementary Issue*, no. 26, pp. 103–121, 2001.

[14] J. Sidnell, "Conversational turn-taking in a Caribbean English Creole," *Journal of Pragmatics*, vol. 33, no. 8, pp. 1263–1290, 2001.

[15] G. F. Simons and C. D. Fennig. (2017) Ethnologue: Languages of the World. Dallas, Texas.

[16] F. Katamba, "Bantu nominal morphology," in *The Bantu languages*, D. Nurse and G. Philippson, Eds. London: Routledge, 2003, pp. 103–120.

[17] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer [computer program]," 2018, version Praat 6.0.37. [Online]. Available: http://www.praat.org/

[18] M. Heldner, "Detection thresholds for gaps, overlaps, and no-gap-no-overlaps," *The Journal of the Acoustical Society of America*, vol. 130, no. 1, pp. 508–513, 2011.