# A cross-linguistic study on turn-taking and temporal alignment in verbal interaction

*Spyros Kousidis*[1], *David Schlangen*[1], *Stavros Skopeteas*[2]

[1]Dialogue Systems Group, Bielefeld University, Germany
[2] General Linguistics Group, Bielefeld University, Germany
spyros.kousidis@uni-bielefeld.de

## Abstract

That speakers take turns in interaction is a fundamental fact across languages and speaker communities. How this taking of turns is organised is less clearly established. We have looked at interactions recorded in the field using the same task, in a set of three genetically and regionally diverse languages: Georgian, Cabécar, and Fongbe. As in previous studies, we find evidence for avoidance of gaps and overlaps in floor transitions in all languages, but also find contrasting differences between them on these features. Further, we observe that interlocutors align on these temporal features in all three languages. (We show this by correlating speaker averages of temporal features, which has been done before, and further ground it by ruling out potential alternative explanations, which is novel and a minor methodological contribution.) The universality of smooth turn-taking and alignment despite potentially relevant grammatical differences suggests that the different resources that each of these languages make available are nevertheless used to achieve the same effects. This finding has potential consequences both from a theoretical point of view as well as for modeling such phenomena in conversational agents.

**Index Terms**: Alignment, Turn-taking, Cross-linguistic

## 1. Introduction

As dialogue unfolds in time, participants take turns in speaking (to a first approximation at least). The resulting temporal [1], rhythmic [2, 3], or durational [4] properties of dialogue speech can be studied for their involvement in two related phenomena: first, they are directly influenced by the pressures of the "cornerstone of conversation" [5], the taking of turns [6], and hence allow inferences about the mechanism that underlies it. Secondly, these properties have often been studied for their role in the phenomenon of speakers becoming "more alike" over the course of a conversation, variously called *synchrony* [7], *accommodation* [8], or *alignment* [9], and observed in different modalities (speech, prosody, posture, gesture and gaze) [10].

The typically studied temporal features are (a) *gaps*, silent time-intervals between speech segments produced by different interlocutors, (b) *pauses*, silent intervals between speech segments of one interlocutor, and *overlaps*, time intevals during which interlocutors speak simultaneously [11]. The study of these features from shallow-annotated corpora offers some insight into the mechanisms of turn-taking [4, 12]. Evidence of temporal alignment comes from observed correlation of per speaker averages of gap, pause and overlap durations over a single interaction in a set of several interactions [4, 10]. However, such evidence is insufficient, as it provides no insights to variation within a single interaction, or to what factors affect temporal alignment in general. "Sliding window" [13, 14, 15] and "beginning vs end" [1] approaches to quantifying temporal alignment in single interactions, show alignment only for some dialogues, while the the effectiveness of these methods is questionable: the width of sliding windows, as well as the division of "early" vs "late" in an interaction are largely arbitrary choices. Thus, the description of temporal alignment phenomena remains an open problem [10].

A less well explored but promising dimension of study is that of cross-linguistic comparison of these phenomena: temporal properties of dialogue that persist across languages could be central in the description of a universal turn-taking mechanism; and temporal alignment, a meta-linguistic property of dialogue speech, is a promising candidate for being a true linguistic universal [16]. [12] investigated temporal phenomena in a diverse set of 10 languages and found roughly similar preferences for "avoiding gaps and overlaps". [3] studied temporal features in 100 dialogues between both native and non-native speakers in Japanese, proposing "simultaneous talking", in contrast to the traditional turn-taking view. [17] investigated gaps on L2 (Korean learners of English) as a measure of fluency. [1] explored the effect of familiarity in temporal alignment in Hungarian, while [18] correlated gap duration with assessment of agreement by interlocutors in Italian, Japanese and American English. Other languages in which temporal phenomena have been studied include Dutch [4, 27] and Swedish [11, 14, 15]. A crucial question is whether temporal phenomena are influenced by properties of linguistic structure (e.g. grammatical properties) or not. For instance, languages with rigid word order provide cues for end-of-utterance prediction [5, 19] (henceforth EUP) that are not available in languages with free word order. Similarly, there is more flexibility in marking phrase boundaries prosodically in intonational languages than in tonal languages. However, it is not clear what effect end-of-utterance prediction has on the exchange of turns.

This paper presents an exploratory study of temporal features using a typologically and culturally diverse set of three languages that is well outside the set of languages studied before. Specifically, we look at floor transitions and evidence for temporal alignment, builing upon and extending previously used methods. We find similarities across languages as well as differences; as we discuss, these findings offer a good starting point for future cross-linguistic investigations into the workings of the interaction-control mechanisms in human dialogue.

## 2. Material

In order to examine the impact of typological properties on the turn-taking latencies we examined a corpus of interactions

in three genetically and geographically remote languages: (a) *Cabécar*, a Chibchan language spoken by 8650 speakers in Costa Rica (according to the 2000 census). The basic word order on this language is SOVX, where the postverbal domain hosts adverbs and prepositional phrases (=X) [20]. Cabécar utterances exhibit final lowering in declaratives and boundary tones associated with final and non-final utterances. (b) Fongbe, a Gbe language spoken by 1,4 million speakers mainly in Benin (a smaller population is living in Togo). Fongbe has a very rigid SVO order. Tonal events are lexically specified, i.e. each syllable is associated with a rising or a falling contour or a mid level tone [21]. There are no instrumental phonetic studies on the intonation of Fongbe; our data suggests that tonal downstep is potentially a cue for EUP. However, given the fact that tonal realisation is lexically specified, we assume that such prosodic cues are less reliable than in an intonational language. (c) Georgian, a Kartvelian language spoken by 3,9 million speakers in Georgia and further populations in several targets of emigration around the world. Georgian is known as a language with very flexible word order: though SOV is considered to be the basic order in this language [22], all permutations of the three major constituents (S, O, V) occur very frequently in the corpus [23]. Georgian is an intonational language employing boundary tones for the indication of intonational phrase boundaries and other devices that are relevant for EUP [24]. The basic syntactic and prosodic properties of the languages in our sample are summarised in Table 1. The straightforward implication of the grammatical observations is that EUP should be more reliable in Cabécar, a language with rigid word order and rich intonational cues for the prediction of the end of the utterance. End of utterance is less reliably predictable through the syntax in Georgian and through intonation in Fongbe.

| | Fongbe | Cabécar | Georgian |
|---|---|---|---|
| syntactic cues | + | + | - |
| tonal cues | - | + | + |

Table 1: *Typological properties relevant for EUP*

The three sub-corpora were collected in the field under identical settings. The main idea of the corpus is to simulate the partition of common ground in a natural conversation: both interlocutors share common access to a subset of the relevant information while they have exclusive access to other subsets of it. In the experimental manipulation, participants were presented with different sets of parts (only some parts are shared between the sets) from a short video (ca. 3 min). After watching the parts separately from each other, the subjects were instructed to make a short conversation in order to figure out what happened in the presented story. The field sessions were video-recorded with a Panasonic Full HD-Camcorder, HDC-SD707EG-K in MPEG-4. Sound recordings were made with two DAT-recorders (Olympus LS-13) in WAV files at 16 bit, 44.1 KHz. The corpus studied in this paper contains: 19 dialogues (58 min) in Cabecar (7 pairs), 31 dialogues (119 min) in Fongbe (7 pairs), and 24 dialogues (60 min) in Georgian (8 pairs).

## 3. Methods

The collected audio files were analysed semi-automatically using the Praat [25] software and following the methodology in [13]: First, speech-silence segmentation was performed using

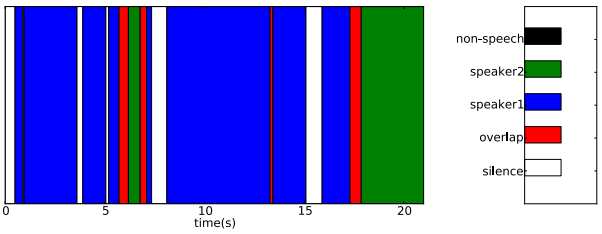


Figure 1: *Schematic of dialogue turn succession (part of dialogue shown)*

the built-in function of the Praat software, which requires setting intensity and duration thresholds. The intensity threshold was set by trial and error depending on the signal-to-noise ratio in the recordings. The minimum durations for silences and speech intervals were set to 50 ms each. While these values are too short according to [26], in our experience they ensure that some valid intervals are not mis-segmented. The algorithm typically mis-segments stop consonants at word final positions, which may erroneously shorten single word utterances below the threshold. Conversely, silences maybe shortened due to environment noise in the recordings. The recordings in this corpus are indeed noisy, and the amount of cross-talk in the audio channels prohibits any automatic separation of the speakers. In a second step, human annotators processed the automatically produced segmentations. The annotation task included manually correcting segmentation boundaries, assigning speech intervals to speakers, and marking non-speech intervals and other parts appropriately (e.g. the experimenter or other people are sometimes heard talking), in order to be excluded from the analysis. Annotators could use the videos in order to assign the correct speaker to an interval in ambiguous cases (e.g. same gender speakers with similar voices). A representation that can easily be derived from these annotations is a *timeline* or *chronograph* which shows the alteration of turns between the two speakers, as shown in Figure 1.

| Abbr. | Description |
|---|---|
| AFT | Average FTO length |
| AGL | Average gap length |
| OVL | Average overlap length |
| APL | Average pause length |
| OFR | Overlap frequency |
| JAT | Joint active time |
| FTC / FTR | Floor transfer count / rate |

Table 2: *Global durational measures and abbreviations*

Intervals of pauses, gaps and overlaps were identified automatically from these timelines, by considering each silent or overlap interval and its two surrounding intervals. If the latter belong to differrent speakers, then this is a floor transfer (FT), either a gap or an overlap, and it is attributed to the speaker that comes *after* the interval in question. If the surrounding intervals belong to the same speaker, then this is "within-turn" event. The overlap intervals of this variety were not considered here. Typically they represent back-channels: short single-word phrases that completely overlap the current speaker, never "holding" the floor. The red bar around 14 seconds in Figure 1 represents such an event, while two long pauses can be seen within the turns of speaker 1 (blue). The measures shown in Table 2 can hence

be computed directly from the timelines. The top four are all averages of a temporal feature over a dialogue or part thereof. [11, 12] considered gaps and overlaps jointly as one distribution, called a *floor transfer object* (FTO) [27], with overlap durations given negative signs. Overlap frequency (OFR) is simply the percentage of all floor transfers that are overlaps.

Previous research [4, 10] speculated that correlation of temporal features among interlocutors may be the result of interactivity, liveliness or engagement in the interaction, rather than of alignment. In order to quantify such factors, [2] used the ratio of vocalisation over silence, while [28] proposed *joint active time* (JAT), which is the percentage of time that any vocalisation occurs, by either or both parties. The study presented here also considered the count and frequency of floor transfers (FTC and FTR, respectively) as a measure of interactivity and explored the effects of (intra-turn) pause length on the FTO duration.

Statistical analysis (e.g. averaging) of durational features may be affected by outliers (such as long silences that were caused by possibly undocumented factors), leading to biased results. [2, 3, 4] used a log transformation of the durational features (which exhibit exponential distributions in raw form), which reduces the effect of outliers, but [11] showed that this can mislead interpretation of the distribution shapes. Other common manipulation methods include identifying and excluding singular outliers, taking the median rather than the arithmetic mean of a feature, or setting thresholds, reducing the range of values [10]. We follow the latter method, setting the valid range at 2.5 standard deviations from the mean.

## 4. Results and discussion

The distributions of floor transfers are shown in Figure 2. Georgian has the smallest average FTO length (mean -96 ms, median 25 ms ) and is characterised by long overlaps. The amount of transfers that are overlaps is 48%. The situation is very different in Cabécar: the mean floor transfer length is 332 ms (median 229 ms), and only 23% of floor transfers are overlaps. Fongbe has the narrowest and most balanced curve, with a positive average floor transfer length of 91 ms (median 91 ms) and 64% of floor transfers on the positive side (gaps). Compared to the other two languages, the percentage of floor transfers that are overlaps in Georgian (51%) seems indeed very high. Two-tailed *Welch's t-tests* confirmed that the FTO means of any two languages are significantly different at 99.9% probability.

These findings contradict our intuitive hypothesis based on the typological properties of these languages: It was expected that Cabécar, with both grammatical and prosodic EUP cues at the disposal of interlocutors would allow for more efficient turn-taking than the other two languages in the set. This is not reflected in the results, in which speakers of Cabécar exhibit the longest floor transfers. Conversely, the unpredictable word order in Georgian does not appear to inhibit swift turn-taking. We can hypothesise that these cross-linguistic differences in the distribution of FTO duration cannot be accounted for through the grammatical differences; they are rather due to cultural conventions in the turn taking behaviour. Despite these differences in the distribution of the FTO for the three languages, we verify the findings of [12] that speakers try to minimise gap and overlap durations (the overall average durations are very close to zero) and that these durations are jointly distributed around one positive peak, for all three languages.

Next, we looked into temporal alignment among interlocutors in each language. Table 3 shows how averages of temporal features are correlated between the participants in an interaction (calculated across all dialogues in each sub-corpus). AFT, the average of the joint ditribution, is correlated in all three languages, while weaker correlations are found for the other measures and only in some languages. An explanation of this may be that overlaps and gaps, which together constitute the floor transfer object FTO, have different pragmatic functions in discourse and that, in some languages, these functions are more clearly separated. E.g. OFR is correlated in Georgian (which has the most overlaps) and AGL in Cabécar, which has the most (and longest) gaps. Similarly, correlations of both average gap and overlap length in Fongbe seem consistent with the symmetric distribution of FTO. On the other hand, average pause length shows weaker or no correlation (not shown), which is a strong indication that temporal alignment phenomena are mostly related to (or are more evident at) floor transfers.

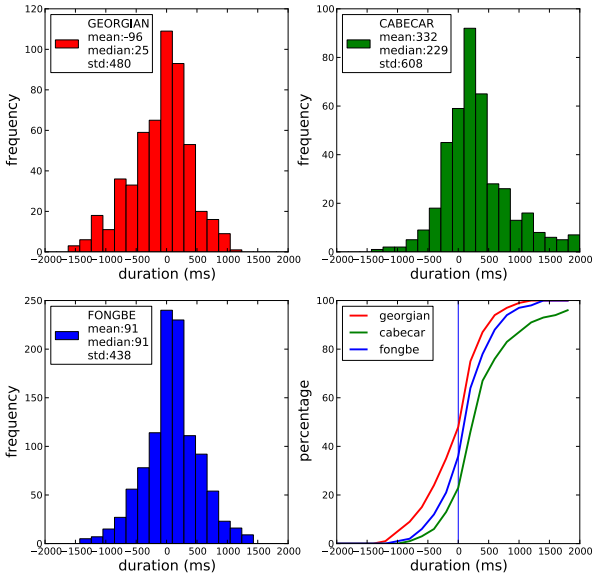


Figure 2: *Histograms of FTO for Georgian, Cabécar and Fongbe, and cumulative FTO distributions for all three languages.*

| Language | Georgian | | Fongbe | | Cabécar | |
|---|---|---|---|---|---|---|
| Feature 1 | r | p | r | p | r | p |
| AFT | 0.69 | 0.001 | 0.59 | 0.001 | 0.53 | 0.034 |
| AGL | - | - | 0.39 | 0.029 | 0.49 | 0.001 |
| OVL | - | - | 0.51 | 0.015 | - | - |
| OFR | 0.48 | 0.017 | - | - | - | - |

Table 3: *Correlations of speaker A vs speaker B across all dialogues per language*

In order to rule out alternative explanations of the observed inter-speaker correlations, we also explore the relationship between the variables shown in Table 4 and the FTO. Joint active time (JAT) is strongly correlated with FTO in all three languages, which verifies the findings of [28] for English and [10] for Japanese: more engaged dialogues with less overall silence lead to "compressed" FTO. This can be seen in Figure 3, showing data from 30 dialogues in Fongbe, for which both JAT and floor transfer rate (FTR) are correlated with FTO, thus most of the bigger, darker circles (coding FTR and JAT, respectively)

are located towards the lower left part of the Figure.

| Language | Georgian | | Fongbe | | Cabécar | |
|---|---|---|---|---|---|---|
| Variable | r | p | r | p | r | p |
| APL | 0.29 | 0.06 | 0.32 | 0.03 | 0.39 | 0.033 |
| APLI | 0.38 | 0.007 | - | - | 0.58 | 0.001 |
| FTR | - | - | -0.49 | 0.018 | - | - |
| JAT | -0.76 | 0.001 | -0.71 | 0.001 | -0.82 | 0.001 |

Table 4: *Predictor variables for FTO across languages (APLI = APL of Interlocutor)*

Intra-sentence pause (APL) has a weak effect on FTO in all three languages, except in Cabécar. Interestingly, a speaker's average floor transfer time is correlated both to one's own average pause length as well as that of the interlocutor's, with the latter effect being more evident in the cofficients. This finding points towards presence of active listening [29], which is the process of monitoring, signaling and contributing while not holding the floor, so that the transition is as smooth and similar to the interlocutor's tempo as possible.

However, when using these variables to *predict* the FTO, by computing them before each FTO instance and using various window lengths, from previous utterance to entire dialogue history, the situation changes radically: none of these variables are good predictors of the FTO, showing zero or very weak correlations ($r < 0.2$). This indicates that these variables are actually side-effects of temporal alignment, or parallel phenomena with little bearing on temporal alignment among speakers. JAT "compresses" FTO, reducing its variance, but does not *cause* the speakers to become aligned. In other words, high engagement is *not* a common external cause that makes the FTOs of the two speakers appear similar. If that were the case, JAT would be a good predictor variable. Similarly FTC and FTR constrain FTO duration but again have minimal predicting power. In the case of APL, we may be observing a rhythmic entrainment phenomenon, sort of a side-effect of alignment at TRPs [6], which pushes interactants to unknowningly adopt a temporal structure in their own speech (rather than the opposite).

Therefore, while these variables have a global effect on FTO, they are not the causes of FTO alignment. We can speculate, that what determines the duration of each FTO instance are grammatical, syntactic, prosodic and non-verbal cues (e.g. gaze), on which native speakers are trained linguistically and culturally. During actual interactions interactants align on these cues, leading to the macroscopically observable correlation in the average duration of FTO among them. For the same reason as all the above, the FTO of the interlocutor is *not* a good predictor of the FTO of a speaker. Globally they are aligned because interlocutors align on the cues, but locally this has no bearing. That is why sliding window methods are successful only if for some reason the discourse leads to semantically or pragmatically similar TRPs, and thus synchronously similar FTOs.

Our findings are similar to [12], which showed that there are crucial cross-linguistic differences in response latencies in conversation that cannot be accounted for by structural differences between the object languages (in particular difference between speech-act indicators that appear early or late in the utterance). Our results here also show glaring differences in FTO duration across languages but do not confirm the pattern predicted by their basic grammatical properties. These findings open the question of whether grammatical properties of languages are relevent for temporal phenomena at turn exchanges;

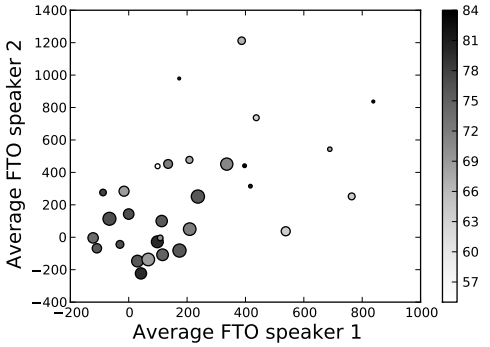


Figure 3: *Scatter plot of Mean FTO duration of interlocutors in 30 Fongbe dialogues, point size indicates FTR, shading shows JAT*

the answer to this question is related with a fundamental issue, namely whether (or how much) cues for EUP influence turn-taking decisions. It is clear that EUP is only a part of the prediction of TRPs, which, in turn, is only one of the factors involved in the observed turn-taking latencies. In comparing different languages, culturally determined properties of speech tempo may have a stronger impact on temporal alignment than the structural properties determining EUP.

## 5. Conclusions

We have expored temporal phenomena in a linguistically and culturally diverse language set. Despite the contrasting language structures and theorised EUP cues, latency distributions of similar shape and evidence of temporal alignment among interlocutors are found in all three languages. The exploration of several factors that are globally correlated to the FTO supported the view that temporal alignment is not a by-product of simple, mechanical phenomena; it may rather be the result of fluency in perceiving relevant cues and identification with a specific culture. In the future, we shall study these cues (grammar, typology, semantics) and cultural factors in order to show how alignment occurs on these, leading to the macroscopically observable corellation of FTO durations. Therefore, more cross-linguistic research is needed in order to identify the components that determine turn-taking latencies; the present study is a contribution to this direction.

## 6. Acknowledgements

This research is partly supported by the Deutsche Forschungsgemeinschaft (DFG) in the CRC 673 "Alignment in Communication". The authors would like to thank Bettina Rempel, Larissa Gbegnonvi, Michael Bartholdt and Jens Eckmeier, for annotating the dialogues. The Cabécar data were collected in Ujarrás, Costa Rica, August 2011, by Carolina Pasamonik; Fongbe data were collected by Larissa Gbegnonvi in Kotonou, Benin, September 2011. The Georgian conversations were collected in Tbilisi, September 2012, by Veronika Ries.

## 7. References

[1] Grászi, T. E. and Bata, S., "The effect of familiarization on temporal aspects of turn-taking: A pilot study," Acta Linguistica Hungarica, vol. 57, pp. 307-328, 2010.

[2] Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P. and Stern, D. N., "Rhythms of Dialogue in Infancy:

Coordinated Timing in Development," Monographs of the Society for Research in Child Development, vol. 66, pp. i-149, 2001.

[3] Campbell, N., "Approaches to conversational speech rhythm: Speech activity in two-person telephone dialogues," in Proc XVIth International Congress of the Phonetic Sciences, Saarbrucken, Germany, 2007, pp. 343-348.

[4] Bosch, L. T., Oostdijk, N. and De Ruiter, J. P., "Durational aspects of turn-taking in spontaneous face-to-face and telephone dialogues," in 7th International Conference TSD 2004, Brno, Czech Republic, 2004, pp. 563-570.

[5] De Ruiter, J. P., Mitterer, H. and Enfield, N. J. ,"Projecting the end of a speaker's turn: A cognitive cornerstone of conversation," Language, pp. 515-535, 2006.

[6] ] Sacks, H., Schegloff, E. A. and Jefferson, G. "A Simplest Systematics for the Organization of Turn-Taking for Conversation," Language, vol. 50, pp. 696-735, December 1974.

[7] Cummins, F. "On synchronous speech," Acoustics Research Letters Online, vol. 3, p. 7, 2002.

[8] Giles, H, Mulac,.A., Bradac, J. J. and Johnson, P. "Speech Accomodation Theory: The First Decade and Beyond," in Communication Yearbook 10, M. L. McLaughlin, Ed., ed Newbury Park: SAGE, 1987, pp. 13-48.

[9] Pickering, M. J. and Garrod, S. "Toward a mechanistic psychology of dialogue," Behavioral and Brain Sciences, vol. 27, pp. 169-190, April 2004.

[10] Kousidis, S., "A Study of Accommodation of Prosodic and Temporal Features in Spoken Dialogues in View of Speech Technology Applications," PhD, DIT, Dublin, 2010.

[11] Heldner, M. and Edlund, J., "Pauses, gaps and overlaps in conversations," Journal of Phonetics, vol. 38, pp. 555-568, 2010.

[12] Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T, Hoymann, G., Rossano, F., De Ruiter, J. P., and Yoon, K.-E., "Universals and cultural variation in turn-taking in conversation," Proceedings of the National Academy of Sciences, vol. 106, pp. 10587-10592, 2009.

[13] Kousidis, S. and Dorran, D. , "Monitoring Convergence of Temporal Features in Spontaneous Dialogue Speech," presented at the 1st Young Researchers Workshop on Speech Technology, UCD, Dublin, Ireland, 2009.

[14] Edlund, J., Heldner, M. and Hirschberg, J. ,"Pause and gap length in face-to-face interaction," presented at the InterSpeech 2009, Brighton, UK, 2009.

[15] Fors, K. L., "Pause length variations within and between speakers over time," in Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue, Los Angeles, USA, 2011.

[16] Evans, N. and Levinson, S. C., "The myth of language universals: Language diversity and its importance for cognitive science," Behavioral and Brain sciences, vol. 32, pp. 429-448, 2009.

[17] Ryoo, H.-K., "Inter-turn Gaps in Small Group Discussion Talk among Korean EFL Learners," Secondary English Education, vol. 4, pp. 3-22, 2011.

[18] Roberts, F., Margutti, P., and Takano, S., "Judgments Concerning the Valence of Inter-Turn Silence Across Speakers of American English, Italian, and Japanese," Discourse Processes, 48(5), pp. 331-354, 2011.

[19] Schlangen, D. "From reaction to prediction: Experiments with computational models of turn-taking," in Interspeech 2006, Pittsbutg, Pensylvania, USA, 2006.

[20] Verhoeven, E., "Cabcar. A Chibchan language of Costa Rica". In: Hurch, B., Sakel, J. and Stolz, Th. (eds.), Proceedings of the Conference European Network of Amerindian linguistics (ENAL), 2012.

[21] Lefebvre, C. and Brousseau, A.-M., A grammar of Fongbe. Berlin: Mouton De Gruyter, 2002.

[22] Skopeteas, S. and Fanselow, G., "Focus in Georgian and the expression of contrast". Lingua 120, 1370-1391, 2010.

[23] Apridonidze, Sh., sitqvatganlageba axal kartuli [word order in Modern Georgian]. Tbilisi: Mecniereba, 1986.

[24] Skopeteas, St., Féry, C., Asatiani, R., "Word Order and Intonation in Georgian". Lingua 119, 102-127, 2009.

[25] Boersma, P. and Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.42, retrieved 2 March 2013 from http://www.praat.org/

[26] Heldner, M., "Detection thresholds for gaps, overlaps, and no-gap-no-overlaps," The Journal of the Acoustical Society of America, vol. 130, p. 508, 2011.

[27] Bosch, L. T., Oostdijk, N. and De Ruiter, J. P., "Turn-taking in social talk dialogues: Temporal, formal and functional aspects," in SPECOM, St Petersburg, 2004.

[28] Kousidis, S., Dorran, D. ,McDonnell, C. and Coyle, E. ,"Towards Flexible Representations for Analysis of Accomodation of Temporal Features in Spontaneous Dialogue Speech," presented at the InterSpeech 2009, Brighton, United Kingdom, 2009.

[29] Campbell, N. "Individual traits of speaking style and speech rhythm in a spoken discourse," Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction, pp. 107-120, 2008.