ISSN 2347-1921

Multimodal Analysis of Topic Shift in Persian Dyadic Conversations

Ghazaleh E. Baiat and Istvan Szekrényes
Department of General and Applied Linguistics, University of Debrecen, Hungary
esfandiari.gh@gmail.com
xepenator@gmail.com

ABSTRACT

Conversations play an important role in our daily life. During conversations, we usually talk from topic to topic automatically, smoothly and effortlessly. However, there are sometimes difficulties experienced in establishing and also acknowledging a new topic in an ongoing conversation. This fact shows that there are specific features and mechanisms to both producing and perceiving topic change, which play a prominent role in keeping continuous talk. The present study provides a multimodal analysis of the occurrence of topic shift in dyadic conversations which were recorded both for audio and video. The research was carried out on two stages, studying both visual and prosodic features of topic shift. On the first stage, the actual speaker's gaze movements around topic shift were investigated while on the second stage, prosodic features such as pitch movement and intensity surrounding topic shift were measured. We wanted to know whether any of these features and their combination could be used as a cue to detect topic shift in a conversation. Focusing on detection and analysis of these features could be helpful for a better understanding of human-human as well as human-machine communication.

Indexing terms/Keywords

Dialogue; Topic shift; Prosodic features; Gaze behavior

Academic Discipline And Sub-Disciplines

Linguistics

SUBJECT CLASSIFICATION

Conversation analysis

TYPE (METHOD/APPROACH)

Multimodal approach

1 INTRODUCTION

Conversations are often characterized as having a certain topic or several topics [1]. While moving from one topic to another is found to be very common in conversations, it is not considered as a random happening. Instead, it is found in specific environments and in characteristic ways. Research in conversational analysis suggests that topicality is an achievement of conversationalist, something organized and made observable in a patterned way that can be described [1].

The present study provides a multimodal analysis of the occurrence of topic shift in dyadic conversations. Conversations, being more than simply a sequence of utterances strung together, are indeed considered as having a multimodal process [2]. There are some non-verbal communicative cues such as gestures, eye gaze, postures, head movements and also a range of prosodic cues, such as pitch movement and intensity that lend naturalness to speech. These cues play an important role in human-human as well as in human-machine communication. They help to compensate for many hidden meanings not present on the surface of spoken language [3]. Thus, in order to fully understand, interpret and describe the structure of topic shift in conversation, studying non-verbal elements, both visual and acoustic, is considered to be crucial. Understanding the role of nonverbal behaviors in conveying conversation structure enables improvements in the naturalness of embodied dialogue systems, such as embodied conversational agents [4].

With the aim to identify and describe non-verbal (both visual and acoustic) features of topic shift in a dialogue, the study was carried out on two stages. On the first stage, relevant gaze behavior was studied in relation to occurring topic shifts while on the second stage; the prosodic features of topic shift were investigated within the dialogue. We wanted to know whether any of these features and their combination could be used as a specific cue to detect topic shift in conversation. To the best of our knowledge, having a multimodal approach towards investigating the occurrence of topic shift within conversation has considerable novelty.

2 RELEVENT TERMINOLOGIES

2.1 Defining ‘Topic’ in Spoken Discourse

There are several definitions to “topic” in spoken discourse. Various authors provide different but not contradictory definitions regarding discourse topic. Chafe [5] claims that a discourse topic ‘‘sets a spatial, temporal, or individual framework within which the main predication holds’’ [6].Li and Thompson [7] state that ‘‘the topic is the center of attention; it announces the theme of the discourse’’. Büring [8] regards topic as the shared ‘aboutness’ of a group of utterances; something that is held in common by multiple utterances within a discourse; the utterances may be ‘about’ a referent, a proposition, or some other entity. He describes discourse topic as “being relevant to but not the same as the possible values for a sentence topic, which contrasts with focused or background information in an individual utterance”. According to him discourse topic constrains the directions which a conversation might take. Hence, it can be seen as an active element in the semantic structure of the discourse which helps to shape the discourse by constraining the directions a following utterance may take [8].

All further uses of the term ‘topic’ in this paper will refer to discourse topic defined by Büring: something that is held in common by multiple utterances within a discourse [8], [9].

2.2 Defining “Topic Shift” in Spoken Discourse

Maynard [1] claims that conversations are often described as having more than one topic. Interaction and collaboration among conversationalists is needed to move from one topic to the other (topic shifting) successfully. While topic change is very common in daily conversation, it is not a random happening. According to Maynard [1] topic shifts are utterances which

- are unrelated to the talk in prior turns
- utilize new referents
- implicate and occasion a series of utterances consisting of a different line of talk

2.3 Defining “Prosody” in Spoken Discourse

Prosodic cues such as rhythm, stress and intonation play a crucial role in providing information regarding the structure of the spoken discourse. Considering each cue as a complex perceptual entity, they are expressed using three acoustic parameters: pitch, energy and duration [10], [11].

Speakers use prosody for a number of communicative purposes, including the following: to lend coherence to shared discourse, to indicate how turns by different participants are tied together into a cohesive, jointly assembled text, and also to express their emotional stance towards the topic-in-progress, such as the degree of enthusiasm or interest they feel for the current topic of discussion [12]. Speakers also show acute sensitivity to one another’s prosody in spontaneous dialogue and use prosody as a resource to convey subtle nuances of expression; for example, when conversation is flowing smoothly, there tends to be a regular rhythm of stressed syllables which is maintained across turns by different speakers, and conversely, a breakdown in this rhythm often signals a difficulty or difference of perspective which needs to be negotiated [12]. In our study, prosody will be especially of interest to us in describing its relation to topic shift within discourse.

3 MATERIAL AND METHOD

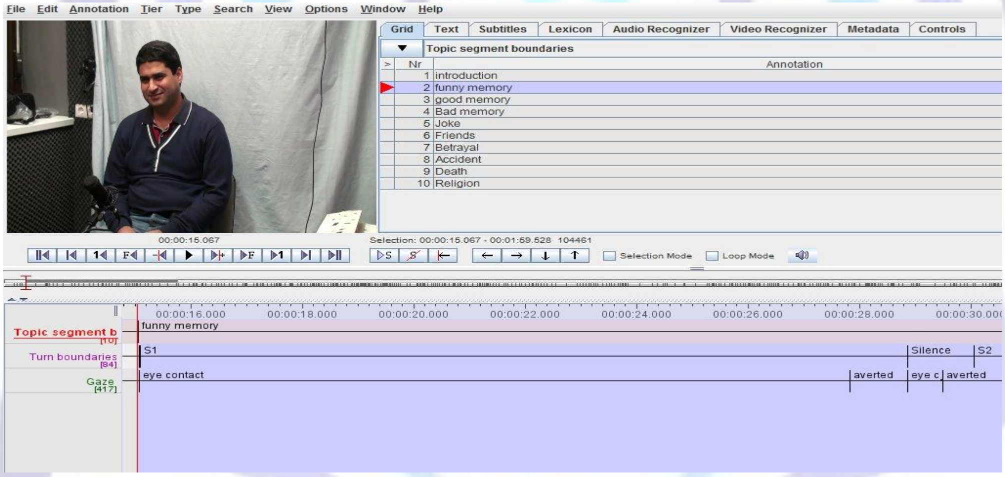
Videotaped dialogues were used as the basis for this study. The recordings, around 40 minutes altogether, were created in a sound-proof studio room. They contained informal dialogue between native Persian-speaking participants. Participants were university students. They were sitting opposite to each other during the conversation. The dialogues were semi-directed by pre-designed topics of talk. The participants were free in handling the topics and they could lead the talk in any desired direction.

3.1 Technical Details

The conversation was recorded with two cardioid microphones (AT2035) and an external sound card to separate the two channels from each other. The recordings were stored in Wave form with 2 channels, 44100 Hz sample rate and 16 bit depth. For annotation and prosodic measurements the tools ELAN [13] and Praat [14] and some algorithms implemented in Praat’s scripting language were used. Our speech rate measurements were obtained using algorithms which were based on the detection of syllable nuclei [15]. For F0 stylization, we used the Prosogram script [16] and our own script named ProsoTool to classify F0 movements near topic boundaries.

3.2 Relevant Gaze Annotation

In order to study the relevant gaze behavior of the speakers, ELAN annotation tool was used. The recordings were subsequently transcribed and annotated using ELAN on three different tiers, each for a main discourse feature: topic boundaries, turn boundaries, and gaze behavior. Figure 1 shows the structure of our annotation scheme.



Nr	Annotation
1	introduction
2	funny memory
3	good memory
4	Bad memory
5	Joke
6	Friends
7	Betrayal
8	Accident
9	Death
10	Religion

Fig 1: Annotation of relevant gaze behavior

A topic segment or boundary was considered as a group of utterances used to convey the purpose of the discourse segment [4], [17]. Thus, the time points at which each of the topics were initiated served as segmentation points. The first and the last word used in each topic block were taken as the initiating and the ending segment of each topic respectively.

Turn boundaries were determined as the point in time in which the start or the end of an utterance co-occurred with a change of speaker, but excluding backchannels [4]. Speech overlaps were also annotated on the turn boundary tier. Also, the beginning and the end of the actual speaker turns were annotated. The “beginning of a turn” was defined as the first word of a new turn. The “end of a turn” was defined as the last word used by the speaker.

Relevant gaze behavior of the speaker was annotated throughout the whole conversation. It was described mainly with two labels: eye-contact and averted. The first was used when the speaker looked towards the hearer while speaking and the second was used when the speaker looked away from the hearer in any other direction.

3.3 Prosodic Annotation

For prosodic analysis of topic shift, Praat was used. Accordingly, the annotations of audio recordings were also implemented with the same tool. On the first tier, we annotated the topic boundaries in the recording.

The next step was to figure out which segmentation principle would be suitable for studying prosodic features of these topic boundaries. In a former publication by Nakajima and Allen [18], the following segmentation principles were mentioned:

- *Pragmatic principle*, where each utterance unit (UU) corresponds to a basic speech act.
- *Grammatical principle*, where UU boundary is set wherever a period can be placed.
- *Conversational principle*, UU boundary should be placed wherever the speaker changes.
- *Prosodic principle*, whenever a medium length or longer pause (750 msec) is inserted between two phrases.

We selected the conversational principle form the above listed options, because in most of the cases in our recordings, topic shift boundaries overlap with turn boundaries. So the whole dialogue was segmented based on speaker change. Three kinds of labels were used in this turn boundary annotation tier: “speaker 1” (SP1), “speaker 2” (SP2) and “overlapping speech” (OS) depending on the actual speaker of the dialogue. Using this segmentation principle, we could investigate the difference between two kinds of speaker-change: the one indicating a topic change and the one not indicating it. During prosodic analysis, segmentation blocks were classified according to their positions relative to the topic boundaries, distinguishing four states: beginning (UU is located at the beginning of a new topic block), ending (UU is located at the end of a topic block), changing (topic shift within an UU) and ongoing (UU is somewhere inside the topic block) part of the topic.

We also used the prosodic principle of segmentation for analyzing the ratio of silences near topic shift boundaries [18].This segmentation was performed in an automatic way. There is a function in Praat which can segment speech signal into “sounding” and “silence” blocks using intensity data for silence detection. In addition to using this function we also inserted the labeling results of ProsoTool into a separated annotation tier. Figure 2 shows the final structure of our annotation scheme.

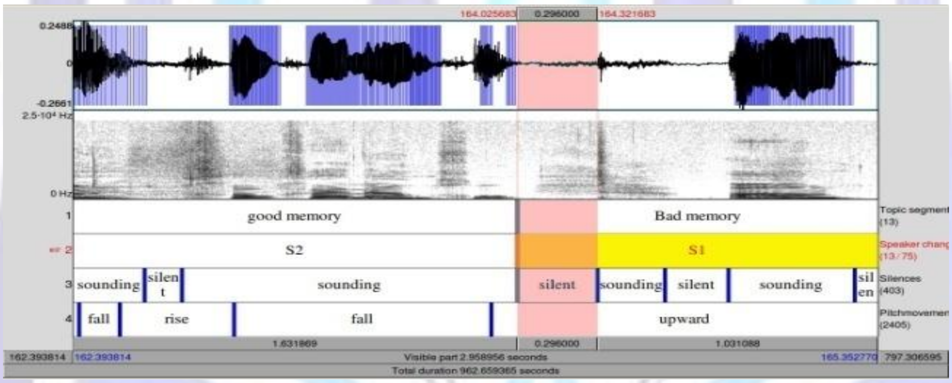


Fig 2: Annotation of audio recordings

4 RESULTS AND DISCUSSION

4.1 Gaze Behavior and Topic Shift

Using data from our dialogues, gaze behavior related to both topic shift and speaker change was investigated. A total 40 minutes of speech was annotated and analyzed in ELAN. Data contained 12 discourse segments (topic units) which constituted 51 turn boundaries (instances of speaker change) altogether. On average, 632 gaze movements were annotated throughout the dialogue having a mean average of 0.26 relevant gaze movements per second. 54% of the gaze movements were annotated as being averted while 46% of the cases were considered as showing eye-contact.

Table 1 summarizes the data related to gaze behavior at the beginning and at the end of topics or, in other words, where topic shifts take place. By “topic beginning” we mean the first word uttered related to the new topic and “topic ending” was marked based on the last word used by the speaker related to that same topic. As it can be seen, in 100% of the cases topic initiation was marked by having an averted eye gaze at the beginning of the utterance followed by eye contact. Regarding topic endings, in most of the cases, 89%, an eye-contact was present between the speakers.

Table 1. Gaze behaviour around topic boundaries

Gaze direction	Topic beginning	Topic ending
Averted	100%	11%
eye-contact	-	89%

Based on our data, the observed pattern showed that whenever a speaker initiated a new topic in the course of conversation, at the very beginning, he tended to look away from his interlocutor. The reason for this kind of behavior (not having direct eye contact) might be related to the speaker wanting to decrease the cognitive load and be more focused on initiating the new topic. Looking at the data we see that the same participant looked directly at his interlocutor while ending the current topic of talk. This can be due to turn-taking mechanisms, the speaker, having eye contact, wants to hand over the floor to the other participant. We could also observe an interaction between topic shifts and turn-taking, topic shifts were more likely to co-occur with turn changes rather than within turns.

Table 2 summarizes the data related to gaze behavior at the beginning and at the end of turn boundaries. Our data showed that in 93% of the cases turn initiation was accompanied with averted gaze while eye contact was only present in 7% of the cases.

Table 2. Gaze behaviour around turn boundaries

Gaze direction	Turn beginning	Turn ending
Averted	93%	31%
eye-contact	7%	69%

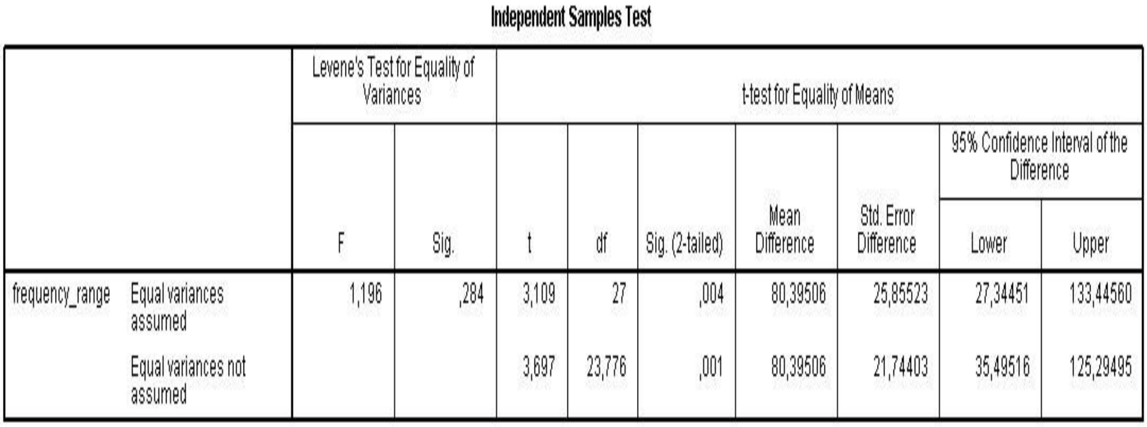
Looking away from one’s interlocutor has been correlated with the beginning of turns also by other studies [4], [19]. From the speaker’s point of view, this look away may prevent an overload of visual and linguistic information [4], [19]. Other studies have shown that eye movements occur primarily at the end of utterances and at grammatical boundaries, and appear to function as synchronization signals. Namely, one may request a response from a listener by looking at the listener, and suppress the listener’s response by looking away [20].

Interestingly, on the other hand, while approaching the end of a turn, speakers tend to use more eye contacts with the hearer. One reason for it might be the intension to offer the floor to the listener [20]. Our study showed that in 69% of the cases turn-ending was associated with having eye-contact but there was also averted gaze present in 31% of the cases.

4.2 Prosody and Topic Shift

In our previous paper [21], we investigated various kinds of prosodic features related to topic shift, describing their differences near the topic boundaries and inside the topic blocks. Due to the lack of the annotation of speaker change that data evaluation had difficulties, because those seemingly significant, descriptive statistical results could only compare the average measurements taken from the whole topic block with the data from topic boundaries. Another issue to remain was that sometimes the results only reflected the different way of speaking of the participants.

According to our new annotation scheme which is based on the conversational principle of segmentation, we could compare the utterance units in different positions (near the topic boundaries and inside the topic) and we could also separate the data coming from separate speakers (S1 and S2). Figure 3 demonstrates the results of T-test related to F0 range of the beginning and the ongoing parts of topic blocks.



Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
frequency_range	Equal variances assumed	1,196	,284	3,109	27	,004	80,39506	25,85523	27,34451	133,44560
	Equal variances not assumed			3,697	23,776	,001	80,39506	21,74403	35,49516	125,29495

Fig 3: Results of T-test related to F0 range of beginning and ongoing parts of topic blocks

By applying this new measuring algorithm to our recordings, we managed to find a statistically significant difference between the F0 ranges of the above mentioned parts of topic units. As it is illustrated in Figure 4, the F0 range of utterance units at the beginning of a new topic was found significantly wider than in the other cases (inside the topic).

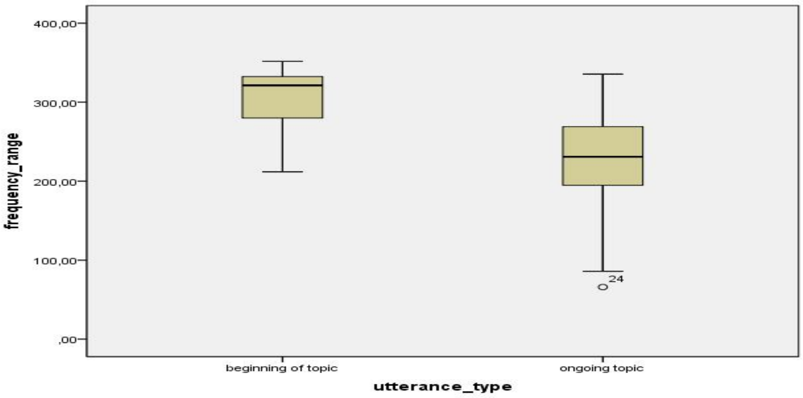


Fig 4: F0 range of beginning and ongoing parts of topic units

5 SUMMARY

The aim of this pilot study was to provide a multimodal analysis of the occurrence of topic shift in a dialogue. In order to do so, gaze behaviour of the actual speaker on one hand, and prosodic features such as pitch movement and intensity on the other hand were investigated. Videotaped dialogues were used as the basis for this study. We wanted to know whether any of these features could be used as a specific cue for topic shift detection in a dialogue.

Regarding gaze behavior, having averted gaze was found as an indicator of both topic initiation and turn initiation while results showed that eye contact was more likely to happen when approaching the end of a topic. Additionally, our results showed that in almost all cases new topic initiations coincided with a speaker change. Although many studies [4], [21], have focused on gaze behavior as a cue only to turn organization, based on the present study we could extend its function by defining that gaze behavior can also be used to detect topic boundaries in conversations.

Our prosodic analysis showed that the F0 range of utterance units at the beginning of a new topic was significantly wider than utterance units in other positions. As a result, we could generalize that the observation of a wide F0 range can be a prosodic indicator of topic shift.

Results such as these can help lead researchers to generate a more natural model of human communication. Automatic topic segmentation is considered as a necessary preprocessing step for applications such as information retrieval, anaphora resolution, and summarization [22]. Due to the limited amount of data being analyzed in the study, further research could definitely be beneficial to increase our understanding regarding the nature of topic shift in conversation.

ACKNOWLEDGEMENTS

We would like to thank Dr. Enikő Tóth from the Department of General and Applied Linguistics for her valuable help with the SPSS analysis. Any errors remaining are our own.

REFERENCES

[1] Maynard, D.W. “ Placement of topic changes in conversation”, Semiotics 30 (3/4): 263-290,1980.

[2] Hunyadi, L. “ Multimodal human-computer interaction technologies”, Argumentum 7: 240–260, 2011.

[3] Wharton, T. 2009. Pragmatis and Non-verbal Communication. Cambridge University, Cambridge

[4] Cassell, J., Nakano, I.Y., Bickmore, W.T., Sidner, L. C., Rich, C. 2001. Non-verbal cues for discourse structure. In Proceedings of ACL:114-123. doi:10.3115/1073012.1073028

[5] Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In: Li, Ch. N. (ed) Subject and topic. Academic, New York, pp 25–55.

[6] Taboada, M. and Wiesemann, L. " Subjects and topics in conversation", Journal of Pragmatics 42:1816–1828, 2010.

[7] Li, Ch. N. and Thompson, S. A. 1976. Subject and topic: a new typology of language. In: Li, Ch. N. (ed) Subject and topic. Academic, New York, pp. 457–489.

[8] Büring, D. 1999. Topic. In: Bosch, P., Rob, van., Der, S. (ed) Focus linguistic, cognitive, and computational Perspectives. Cambridge University, Cambridge, pp. 142-165.

[9] Zellers, M. and Post, B. 2009. Fundamental frequency and other prosodic cues to topic structure. In Proceedings of IDP, Paris

[10] Mary, L., and Yegnanarayana, B. "Extraction and representation of prosodic features for language and speaker recognition", Speech Communication 50: 782-796, 2008.

[11] Shriberg, E., Stolcke, A., Hakkani-Tur, D., Tur, G. "Prosody based automatic segmentation of speech into sentences and topics", Speech Communication 32: 127-154, 2000.

[12] Skidmore, D., and Murakami, K., 'How prosody marks shifts in footing in classroom discourse", International Journal of Educational Research 49: 69–77, 2010.

[13] Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. 2006. ELAN: A professional framework for multimodality research. In Proceedings of LREC, Fifth International Conference on Language Resources and Evaluation. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands http://tla.mpi.nl/tools/tla-tools/elan/

[14] Boersma, P. and Weenink, D. 2011. http://www.praat.org

[15] Yong, N. H. D., and Wempe, T. "Praat script to detect syllable nuclei and measure speech rate automatically", Behavior Research Methods 41 (2): 385-390, 2009.

[16] Alessandro, C.'d., and Mertens, P. 2004. Prosogram: semi-automatic transcription of prosody based on a tonal perception model. In Proceedings of the 2nd International Conference of Speech Prosody: 23-26.

[17] Grosz, B., and Sidner, C. "Attention, intentions, and the structure of discourse", Computational Linguistics 12: 175-204, 1986.

[18] Nakajima, S., and Allen, J. F. "A study on prosody and discourse structure in cooperative dialogues", Phonetica 50: 197-210, 1993.

[19] Chovil, N. "Discourse-oriented facial displays in conversation", Research on Language and Social Interaction 25:163-194, 1992.

[20] Cassell, J., Torres, E. O., Prevost, S. "Turn taking vs. discourse Structure: how best to model multimodal conversation", Machine Conversations: 143-154, 1998.

[21] Esfandiari-Baiat, Gh., Szekrényes, I. Topic Change Detection Based on Prosodic Cues in Unimodal Setting. In Proceedings of CogInfoCom conference, Kosice, Slovakia

[22] Levow, G.A., 2004. Assessing prosodic and text features for segmentation of mandarin broadcast news. In Proceedings of HLT-NAACL-Short:137-140.