The Prosody of Discourse Structure and Content in the Production of Persian EFL Learners

Vahid Sadeghi
Associate Professor
Imam Khomeini International University
vsadeghi5603@gmail.com

Niloofar Mansoory Harehdasht
Assistant Professor
Payame Noor University
nmansoory@gmail.com

Abstract

The present research addressed the prosodic realization of global and local text structure and content in the spoken discourse data produced by Persian EFL learners. Two newspaper articles were analyzed using Rhetorical Structure Theory. Based on these analyses, the global structure in terms of hierarchical level, the local structure in terms of the relative importance of text segments and the rhetorical relations between text segments were identified. The texts were read aloud by 18 high-proficient Persian learners of English. We measured pause durations preceding segments, F0-maxima (as a correlate of pitch range) and speech rates of the segments. Results suggested that speakers give prosodic indications about hierarchical level by means of variations in pause duration but not pitch range or speech rate. Furthermore, it was found that speakers articulate causally related segments with shorter in between pauses and at faster rate than non-causally related segments. However, they did not vary any of the prosodic measures to distinguish between important vs. unimportant segments. Overall, the results suggest that (1) variations in pause duration and speech rate are not used systematically by Persian EFL speakers as fully structured cuing devices to indicate organization of spoken discourse, importance of sentences and meaning relations between sentences; (2) pitch change structuring as a cue to discourse prosodic prominence is completely absent in Persian learners' text production. The results are not consistent with earlier findings of prosodic realization of text structure and content in English native speakers' discourse data, but are in line with those obtained for non-native spoken discourse.

Keywords: Global and Local Text Structure, Rhetorical Relation, Rhetorical Structure Theory, Hierarchical Levels, Prosodic Realization

Received: January 2016; Accepted: March 2017

1. Introduction

Research on text prosody focuses on the prosodic characteristics of clauses and sentences in relation to their position and function in text. In recent years, there has been accumulating evidence that speech texts are produced and interpreted by native speakers and hearers using phonetic cues that appear at different boundaries in the text structure. It has been suggested that speakers give prosodic indications about hierarchical level by means of variations in pause duration and pitch range; in particular, the higher the segments are connected in the text structure, the longer the preceding pauses and the higher the F0-maxima are realized (den Ouden, Noordman & Terken, 2009; den Ouden & Terken, 2001,; Pickering, 2004; Schilperoord, 1996). For example, previous studies by Hirschberg and Grosz (1992), Hirschberg and Nakatani (1996), Lehiste (1975), Silverman (1987), Swerts (1997), Thorsen and Gronnum (1985) and Yule (1980) have all indicated that first sentences of paragraphs have longer preceding pauses and higher pitch range than sentences within paragraphs, and that parenthetical and final sentences of paragraphs are articulated with lower pitch range and at faster rate than sentences at other locations in the text. Further research also investigated the prosodic realization of content relations between sentences in a text. For example, van Donzel (1999) applied the content-related distinction between new, inferable and evoked information as proposed by Prince (1981), and found that there is a systematic relation between prosodic composition and content of a text: new information is realized consistently with a more prominent prosody than the other types of information. The research conducted by Noordman, Dassen, Swerts and Terken (1999) provided empirical evidence that the heights of fundamental frequency peaks as well as the durations of pauses preceding sentences gradually decrease as hierarchical

levels decrease, where high levels are associated with more important information. In a more recent study, den Ouden et al. (2009) examined the organizational and content-related aspects of text, and found that speakers articulate important segments more slowly than unimportant segments, and that they read aloud causally related segments with shorter in-between pauses and at faster rate than non-causally related segments.

Studies of L2 discourse that have investigated the role of prosodic structure in signaling global and local organization of a text have demonstrated crucial differences in the use of prosodic cues by L2 speakers (Anderson-Hsieh, Johnson, & Koehler, 1992; Pickering, 2001; Tyler, Jefferies and Davies, 1988; Tyler & Davies, 1990; Wennerstrom, 1994). In an experimental study, Wennerstrom (1994) examined the spoken discourse data of 30 intermediate learners from three L1 backgrounds performing a read-aloud task, and found that while Spanish L2 speakers and a native speaker control group marked a new topic with a high pitch onset, Thai and Japanese L2 learners demonstrated no paragraph initial pitch change. In addition to pitch variation, research on fluency (i.e., pause structure and hesitation phenomena) in L2 discourse reveals qualitative differences in both placement and length of pauses as compared to an L1 model. It has been found that pauses in L2 discourse are both longer and more erratic than those in the native discourse data, and that these pauses tend to regularly break up conceptual units (Rounds, 1987). Furthermore, studies by Riggenbach (1991) and Anderson-Hsieh and Venkatagiri (1995) report that there are more non-lexical fillers and unfilled pauses in non-fluent L2 speech, and that long pauses frequently appear within rather than between intonation units.

The present paper studies the prosodic realizations of global and local text structure in the production of Persian EFL learners of English. Using a

theoretical framework for the organization of natural text, this study examines both pause and pitch structure in Persian learners' naturally spoken discourse data. We address the question how the structure and content of a text affect Persian learners' prosodic production when they read aloud the text, and to what extent, the patterns of prosodic features characterizing the global structure, rhetorical relations and the local structure of text in Persian learners' discourse data are different from those in native speakers' data.

2. Theoretical Framework: Rhetorical Structure Theory

The discourse structure variables in this study are derived from a discourse representation constructed using Rhetorical Structure Theory (Mann and Thompson, 1988; henceforth: RST). RST is a descriptive theory for the description and interpretation of text structure and content. The theory distinguishes between various text characteristics: the global structure of text in terms of hierarchical level, the rhetorical relations between the sentences, and the local structure in terms of the relative importance of text segments (den Ouden et al., 2009). The first step in analyzing discourse structure in RST is segmentation. The theory assumes a syntactic criterion for segmentation according to which segments are essentially clauses (a clause is defined as (a part of) a sentence that contains a finite verb). Clausal components such as subjects and complements, and restrictive relative clauses are treated as parts of their respective clauses rather than separate segments (see Table 1). This analytic approach to segmentation is essentially different from those (e.g. Segmented Discourse Representation Theory, or SDRT) in which sentential portions are treated as discourse segments if they serve "a discernible discourse function" (Asher & Lascarides, 2003).

Table 1.*Example text (Source: de Volkskrant, 3 November 2000)*

Segment	
1. The census in China has been extended by five days	15. The employees of the demographic committee now admit that most people did not keep the one child policy.
2. Normally it should have ended last Friday.	16. One of these days even a family with ten children has been found in the region Shanxi.
3. However, millions of people avoided the pollsters.	17. In the opinion of the authorities, the counting of the homeless is not problematic.
4. or refused to open their doors	18. It was said that there would not be many homeless people
5. This boycott was intended to keep secret illegal children or addresses	19. and most of them would have an address in another region.
6. During emergency talks last Friday, the government, i.e., the Chinese cabinet, decided to extend the census.	20. They would be counted at these addresses.
7. An official of the census committee in Beijing said that the committee's employees noticed that it was rather difficult to find active people at home by day or during the evening, but that of course many other people avoided them deliberately.	21. However, how this counting should happen is unclear
	22. Other people argue that their privacy is affected.
8. At least eighty million farmers, and that number could even be two hundred million, have squatted in the cities.	23. These people were found especially in the random sample of 10% of the population which had to answer 49 detailed questions.
9. They had themselves registered at their addresses of origin	24. The remaining 90% only had to answer nineteen general questions.
10. or they were not counted at all	25. People who have not yet been counted are encouraged by
11. Although they were officially assured that the census has nothing to do with the police,	advertisements to report that to the census committee.
12. many people are afraid of reprisals when it is discovered that they do not have residence permits.	26. The people who have avoided meeting the pollsters thus far will probably not answer this call.
13. Also many married people who have more than one child boycotted the census.	27. The whole affair has not been proved to be very helpful to the accuracy of this fifth census in the 51-year-old history of the People's Republic of China.
14. because they were afraid that the committee of birth control would find this out.	

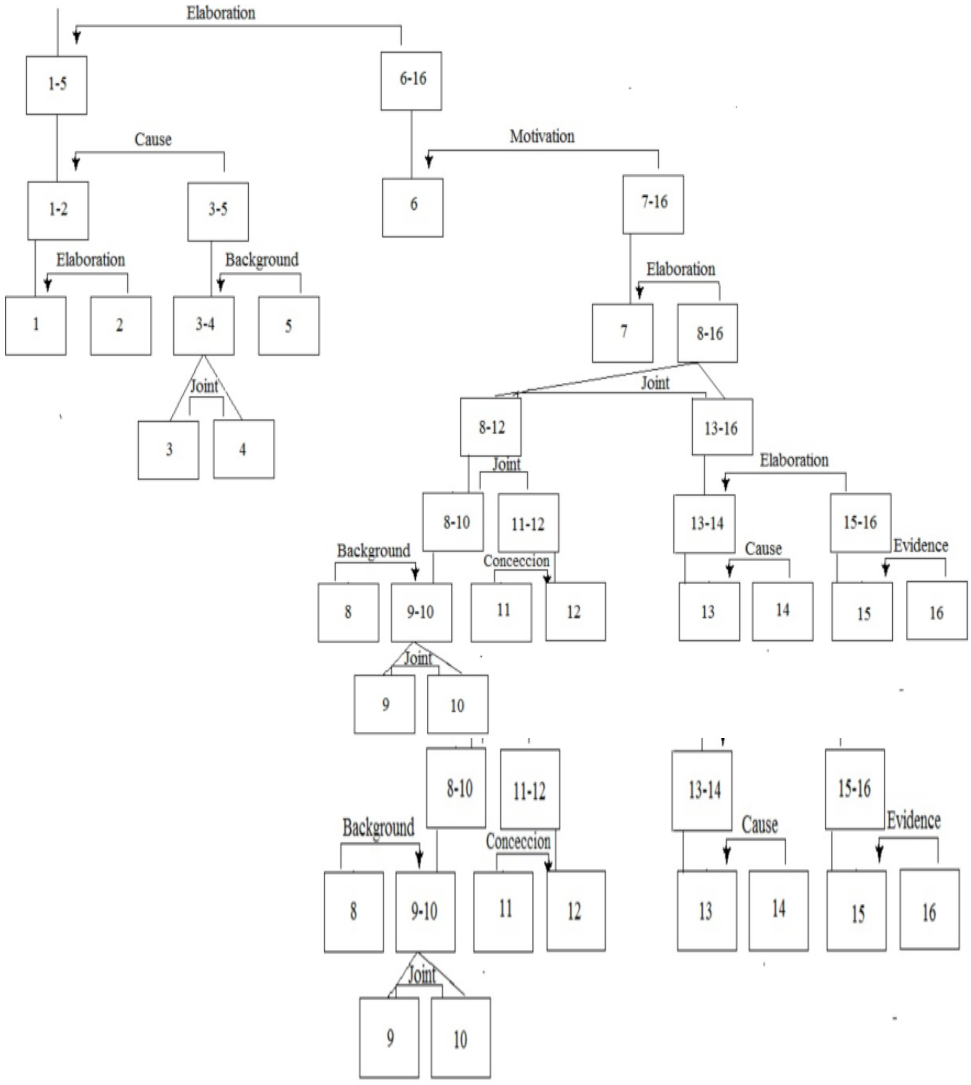

Figure 1. *RST Analysis of Segments 1-16 in Example Text in Table 1*

After segmentation, the analyst identifies rhetorical relations that are inferred to hold between those segments. Finally, the analyst determines the global structure of the text in terms of hierarchical level, showing how

individual segments are related to each other, and how each segment contributes to the overall communicative goal of the text.

A number of rhetorical relations are assumed under RST including Volitional and Non-volitional Cause, Result, Consequence, Background, Motivation, and Elaboration. In each rhetorical relation, one segment is a nucleus, and the other the satellite. The nucleus is the central part of a text span, or the most important part of it; while, the satellite is peripheral or less important, in the sense that a text without satellites can still be understood (den Ouden et al., 2009).

Fig. 1 shows the RST analysis of the hierarchical organization of the segments 1-16 of the example text in Table 1. Rhetorical relations between different parts of the text are indicated by the arrows in the figure. In each rhetorical relation, the point of an arrow indicates a nucleus, while its end indicates a satellite. The 'joint' relation involves two nuclei. Each box represents the segments that form a text span. The relation between text span 1–5 and text span 6–16 is characterized by Elaboration. This means that the content of segments 6–16 is considered to be an elaboration of the content of segments 1–5, which follows from the definition of the Elaboration relation as described by Mann and Thompson (1988). One level lower in the hierarchy, text span 3–5 is characterized as a Volitional Cause of text span 1–2. One level lower, the segments are related to each other by way of an Elaboration: segment 2 elaborates the statement of the nucleus, segment 1. Segment 5 gives background information to segments 3 and 4, based on the relation definition of Background. In this way, all relations between text spans and between the individual segments on the bottom-level are analyzed on the basis of the relation definitions in (Mann & Thompson (1988)).

A text analysis using RST involves top-down and bottom-up analysis at the same time (Bateman & Rondhuis, 1997; den Ouden, 2004; den Ouden et al., 2009; Andreeva, Barry & Koreman, 2016). Following a top-down procedure, the analyst divides the whole text into two large text spans and determines the rhetorical relation between them. These text spans are in turn decomposed into two smaller text spans and the rhetorical relation between these spans is determined, until finally the level of the individual segments is reached. The bottom-up process works the other way around: the analyst relates two individual segments and assigns a relation definition to the pair of segments, thereby creating a text span; this text span is in turn related to another text span and a relation definition is assigned to it, and so on. Both strategies, top-down and bottom-up, are applied simultaneously until all segments of the text are connected in a tree-like structure.

Rhetorical relations in discourse prosody research are often reduced to a smaller, hence manageable, number of content relations (den Ouden, 2004; den Ouden et al., 2009; Sanders, Spooren & Noordman, 1992; Andreeva, Barry & Koreman, 2014). Two most common contrasting pairs of content relations are causal versus non-causal relations and semantic versus pragmatic relations (Sanders et al., 1992). With regard to the first contrast, representing a cause-based rhetorical relation according to Sanders et al. (1992), a non-causal, or additive, relation exists if a conjunction relation can be deduced between two segments, whereas a causal relation exists if a relevant implication relation can be deduced (Sanders et al., 1992). (1) shows an example of a non-causal relation between two segments, while (2) represents an example of a causal relation between two segments.

(1) John stayed at home and his brother went to the party.
(2) John stayed at home, because he wanted to study for his exam.

In (1) the second segment is an addition to the first segment, whereas in (2) the second segment of the pair caused the first segment.

Concerning the semantic versus pragmatic relations, representing a source-based rhetorical relation according to Sanders et al. (1992), a semantic relation exists if the coherence between segments is based on the coherence between the events in the world which are described, whereas a pragmatic relation exists if the coherence between segments is based on the illocutionary meaning of one or both of the segments, for example, when a writer or speaker draws a conclusion (den Ouden et al., 2009; Sweetser (1990). An example of a semantic pair is (3); an example of a pragmatic pair is (4), both cited from Sweetser (1990).

(*3) Anna loves Victor because he reminds her of her first love.*
(4) Anna loves Victor, because she'd never have proofread his thesis otherwise.

In (3) there is a consequence-cause relation of two events in the world, whereas in (4) the second segment can be paraphrased as 'I conclude that she loves him because I know the relevant data'. Sweetser (1990) argues that in a semantically related pair like (3), the consequence in the first segment is presupposed and only the causal relation between both segments is affirmed, whereas in a pragmatic pair like (4) the conclusion in the first segment cannot be presupposed. She assumes therefore that pragmatically related segments require comma intonation, i.e., longer pauses, whereas semantic readings do not (Sweetser, 1990, p. 82).

3. Method

3.1. Materials

Two newspaper reports were selected as the text materials of the study. The mean length of the texts was 33 segments (Text 1 (given in Table 1) included 27 and Text 2, 39 segments). These reports were selected as they represented naturally occurring texts instead of constructed ones, and also as they allowed to distinguish between various text characteristics. The texts were divided into segments (clauses) according to the criteria given by RST. After segmentation, we analyzed the 2 texts using RST. The global structure of the 2 analyses was operationalized in terms of hierarchy; the rhetorical relations between the sentences of the 2 texts were operationalized in terms of causal versus non-causal and semantic versus pragmatic relations; and the local structure of the analyses was operationalized in terms of nuclearity. These operationalisations are explained below.

Following den Ouden et al. (2009), we determined the hierarchical levels of the segments on the basis of the depth of the boundaries between adjacent segments. For each boundary, the superordinate node connecting the two segments adjacent to the boundary was determined, and the number of subordinate nodes including the connecting node itself was counted; the total number of nodes was considered the score of the boundary. Thus, using this procedure, low boundaries received low scores and high boundaries received high scores. For example, in Fig. 2 the boundary between segments 12 and 13 is scored as 5: the superordinate node of segments 12 and 13 dominates seven nodes, two at the left side and two at the right side, and one is added for the connecting node itself. In the same way the boundary between segments 1 and 2 is scored as 1; the boundary between segments 2 and 3 as 4, and so forth.

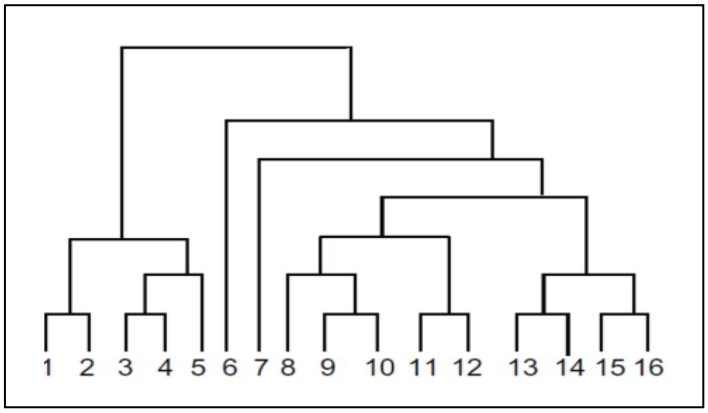

Figure 2. *Bottom-up Representation of Text Analysis in Fig. 1. Numbers Represent the Segments as They Occur in Table 1.*

The relations between segments in the hierarchical structures were then classified in accordance with their associated rhetorical content. For example, the relation between segments 1 and 2 in Fig. 1 was classified as an Elaboration; the relation between segments 4 and 5 was classified as Background. After identifying the rhetorical relations, we classified them in terms of Sanders et al. (1992)'s general contrasting pairs of content relations as causal versus non-causal relations and semantic versus pragmatic relations.

Following the analyses of rhetorical relations, the nuclei and satellites for the boundaries in the hierarchical structures were defined on the basis of the relation definitions. In the RST analysis of the example text in Fig. 1, the nuclei are segments 1, 3, 4, 6, 7, 9, 10, 12, 13 and 15; and the satellites segments 2, 5, 8, 11, 14 and 16. In the 2 RST analyses, nuclei outnumbered satellites because many segments were connected by a Joint, Sequence, or Contrast relation. These relations consisted of two (or more) nuclei.

3.2. Subjects and Data Collection Procedure

The participants were 18 high-proficient Persian EFL learners who were either MA students or MA graduates of English TEFL. None of the participants had any experience being immersed in a native English language environment. To select a homogenous sample group for the research, initially 56 people took a TOEFL English language proficiency test, namely, Oxford Proficiency Test. The test was taken in a quiet room, and the examiner supervised the test administration. The subjects were given one hour to answer 60 questions, and then the answers were analyzed based on the criterion the test had provided. Thus, as determined by the Oxford Proficiency Test, those who scored between 48 to 60 were classified as high-proficient learners (34 people), and eighteen participants were randomly selected from them as the final participants. High-proficient participants were chosen because they were expected to be particularly capable of identifying the discourse structure of a news article, a necessary first step to test the larger question of how prosody correlates with discourse structure.

The participants were instructed to read aloud the text as clearly and naturally as possible. The speakers prepared the reading aloud carefully. They were encouraged to make notes in the text to improve their reading aloud. The preparation was intended to focus the readers' attention on the content and the structure of the text, and to enable them to read it aloud as much as possible in accordance with their mental representation of the text.

The speakers were individually recorded in a quiet room using a digital audio recorder (Sound Blaster X-Fi 5.1) and a Shure directional condenser microphone (SM 58).

3.3. Prosodic Analysis

With respect to previous literature on the prosodic realization of global and local text structure and content in English native speakers' discourse data, we assumed three prosodic features to be systematically correlated with text characteristics, namely pause durations between the RST-defined segments, pitch range and articulation rate of segments. Earlier research has showed that pause durations are shorter, pitch range is lower and articulation rate is faster as the hierarchical levels in the global text structure are lower (for example, Schilperoord, 1996). Concerning the content-based discourse prosody, Sanders and Noordman (2000) have found that sentences that are causally related have faster reading times than sentences that are non-causally related, suggesting that non-causal relations are connected less strongly than causal relations. Similarly, den Ouden et al. (2009), Calhoun (2010) and Braun (2015) have suggested that speakers also need less time to produce these sentences when they read them aloud. The rationale for the prediction is that speakers – as readers – process causally related sentences faster than non-causally related ones and therefore also produce them at a faster rate, or, alternatively, because faster reading more explicitly expresses the relatedness of the causal relation. Concerning the prosodic realization of local text structure in English native spoken discourse, earlier research has revealed that more important information is realized with a more prominent prosody than other types of information. More specifically, it has been found that nuclear segments have longer preceding pauses, higher pitch range, and a slower articulation rate than satellites (den Ouden et al., 2009; Sanders & Noordman, 2000; Schilperoord, 1996; Andreeva et al., 2014, 2016).

Based on what is known about the characteristics of spoken text prosody in English, we were interested to know whether non-native spoken texts produced

by Persian high proficient EFL learners show patterns of prosodic composition that match those found in English native speakers' data. Thus, we measured the same prosodic features as reference parameters to be able to compare our results with those obtained for English native speakers. All acoustic measurements were made using Praat acoustic software (Boersma & Weenink, 2005). For pause duration, the periods of silence at the boundaries between the segments were determined manually by visual inspection of the waveforms and spectrographs, and the duration in between each boundary was determined in milliseconds. Pitch range was measured as the highest F0 value of a segment over the course of its pitch contour (Braun, 2015). During pitch range measurement, we corrected for pitch-measurement errors, like voiced–unvoiced errors as well as F0-maxima associated with final rises (Andreeva et al., 2016). Speech rate was defined as the number of words per second. The number of phonemes was computed automatically in the software using a phoneme detection script command. To account for the reliability of the analyses, all prosodic measurements were double-checked by the first author on different days. Results obtained from different measurements of the same parameters were only slightly different in terms of absolute values.

3.4. Results

3.4.1. Effect of Hierarchy

The graphs in Fig. 3 plot mean pause duration (top left), F0-maximum (top right), and speech rate (the number of words uttered per second) (bottom) for each hierarchical level in global text structure. The results are given only for initial segments. The data for the non-initial segments were not analyzed due to the lack of observations at higher levels 3 and 4. Separate one-way repeated-measures ANOVAs were performed across speakers for the three measures of

prosody; i.e., pause duration, F0 maximum and speech rate, with hierarchy (4 levels) as the independent variable and speaker as the random factor.

The data reveal a consistent effect of hierarchy on pause duration: Durations of pauses increase with hierarchical level. Pause duration is longest for hierarchical level 4 and shortest for Hierarchical level 1, and intermediate for hierarchical levels 2 and 3. The mean differences between the lowest (level 1) and the highest level (level 4) ranged from 462 ms to 856 ms depending on the speaker. The ANOVA for pause duration revealed a main significant effect of hierarchy ($F(3,54)=138.83; \varrho < 0.001$). Durations of pauses differed from each other according to Scheffe post-hoc tests (all pairwise comparisons, $\varrho<0.001$).

However, unlike pause duration, pitch range values do not reveal a consistent effect of hierarchy. Values of F0 maxima (the acoustic parameter for pitch range operationalization) are realized with little between-groups variability within a pitch range of 230-240 Hz irrespective of the size of the hierarchical levels. In addition, as shown in the figure, speakers showed variation in mean F0 points that ran against the predictions of our hypothesis.

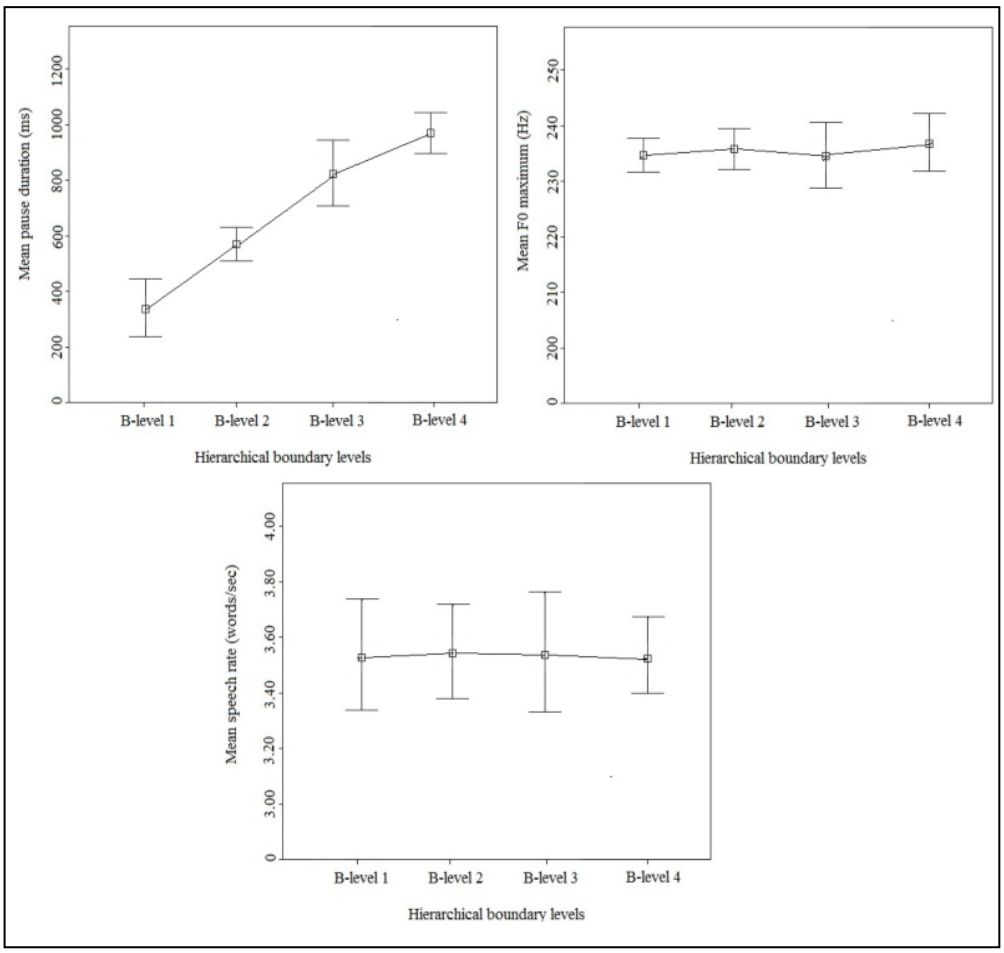

Figure 3.*Mean and Standard Error values of Pause Duration (top left), F0-maximum (top right), and Speech rate (bottom) as a Function of Hierarchical Level in Global Text Structure for all Speakers.*

For example, segments associated with hierarchical level 2 were realized with a higher pitch range value than those associated with hierarchical level 3 (which were expected to be higher in F0 maxima than the lower boundary levels 1 and 2). Thus, when the prosody of boundaries in the hierarchical structure was measured with reference to F0 peak over the course of a segment's pitch

contour, discourse segments preceded by larger boundary sizes exhibited no higher F0 maxima than those preceded by smaller boundaries. As expected, the ANOVA for F0 maxima did not show a significant effect of hierarchy ($F(3,54)= 0.60$, $p=0.439$).

Similar to pitch range, the data for speech rate display no clear effect of hierarchy. In general, speech rate is rather constant across different boundary sizes. Speakers did not display a boosting (or lowering) effect on speech rate as a discourse segment's preceding boundary increased in size (four of the speakers showed some small differences in speech rate as a function of boundary size, with the rate of speech being generally higher for larger boundary levels than for smaller levels. However, even for these speakers, hierarchy did not consistently affect speech rate). As expected, in the ANOVA for speech rate, no effect of hierarchy was found ($F(3,54)=1.33$; $\varrho= 0.27$).

In sum, the results of this section indicate that while pause duration is consistently affected by the size of the hierarchical levels in global text structure, pitch range and speech rate show little variation with hierarchical level. These results are not generally consistent with previous results obtained for English native speakers. As explained earlier, previous research on the prosodic realization of global text structure in English native speakers' discourse data has showed that all the three measures of prosodic realization vary significantly with hierarchical level, with pause duration being longer, pitch range being higher and articulation rate being slower as the hierarchical levels in the global text structure are higher (den Ouden et al., 2009; Schilperoord, 1996).

3.4.2. Effect of Rhetorical Relations

The effect of rhetorical relations on prosody was addressed with respect to both causal and non-causal relations and semantic and pragmatic relations. Both analyses concerned the position of the second segment of the relation, because the relation was assigned to the boundary preceding the second segment of the related pair. The second segment in each rhetoric relation was either initial or non-initial. The data were submitted to separate two-way repeated measures ANOVAs for each of the three measures of prosody; i.e., pause duration, F0 maximum and speech rate, with rhetorical relation and position as the independent variables and speaker as the random factor. Hierarchy was not included as an independent factor, since there were too low frequencies in some levels of the hierarchical structure.

First, we present results for causal and non-causal relations. The graphs in Fig. 4 plot mean pause duration (top left), F0-maximum (top right), and speech rate (bottom) across all speakers as a function of cause-based rhetorical relation (causal vs. non-causal) and position (initial vs. non-initial). The results for pause duration indicate that values for this measure are radically different across the two cause-based rhetorical relations in both the initial and non-initial positions: Pauses are shorter between causally related segments than between non-causally related segments. Pauses are also shorter preceding non-initial segments than initial segments. Results of ANOVA analysis for pause duration showed a significant effect of cause-based rhetorical relation ($F(1, 62)=83.29$; $\varrho<0.001$). The results further revealed a significant effect of position ($F(1,62)=77.16$;$\varrho<0.001$). The interaction between cause-based rhetorical relation and position was not significant ($F<1$).

Unlike pause duration, pitch range was not affected by cause-based rhetorical relation or position of the second segment. The data show that, in

general, values of F0 maxima are rather constant across both cause-based rhetorical relations in initial and non-initial positions. The mean values of F0 peak over the course of a segment's pitch contour, applied to all speakers, were 238 Hz and 232 Hz for causally related segments in initial and non-initial positions, and 237 Hz and 236 Hz for non-causally related segments in initial and non-initial positions, respectively. The results of ANOVA analysis did not reveal an effect of cause-based rhetorical relation ($F(1, 62)=2.89$, $\varrho=0.11$) or position ($F(1, 62)= 0.69$; $\varrho= 0.51$) on F0 maxima.

The results for speech rate reveal a clear effect of cause-based rhetorical relation. Causally related segments are articulated at a faster rate than non-causally related segments, suggesting that speakers produced more phonemes per second when reading causally related segments. However, unlike cause-based rhetorical relation, the position of the second segment seems to have no clear effect on speech rate. The ANOVA revealed a significant main effect of cause-based rhetorical relation ($F(1, 62)=129.49$, $\varrho<0.001$).; however, there was no effect of position and no interaction (both, $F<1$). Next, the three measures of prosody were analyzed with respect to source-based rhetorical relation (including sematic and pragmatic relations) between segments. Table 2 reports mean and standard error values of pause duration (ms), F0-maximum (Hz), and speech rate as a function of source relation and position of the second segment of the pair. The data reveal no clear effects of source-based rhetorical relation or position on F0-maximum or speech rate. Values for these two measures are rather stable across the two different source relations and positions. The ANOVAs for the two measures did not show a significant effect of source relation (F0-maximum: $F(1,42)=0.69$; $\varrho=0.51$; speech rate: $F(1,42)=1.76$, $\varrho=0.23$) or position (F0-maximum: $F(1,42)=2.52$, $\varrho=0.11$; speech rate: $F(1,42)=0.46$, $\varrho=0.49$). Also, there were no significant

interactions (both, F<1). Similarly, the data for pause duration showed no effect of source relation (F(1,42)= 0.44, ϱ=0.5). However, position did affect values for this measure (F(1,42)= 20.10, ϱ<0.001), with initial second segments having longer preceding pauses than non-initial second segments. There was no interaction between source and position (F<1).

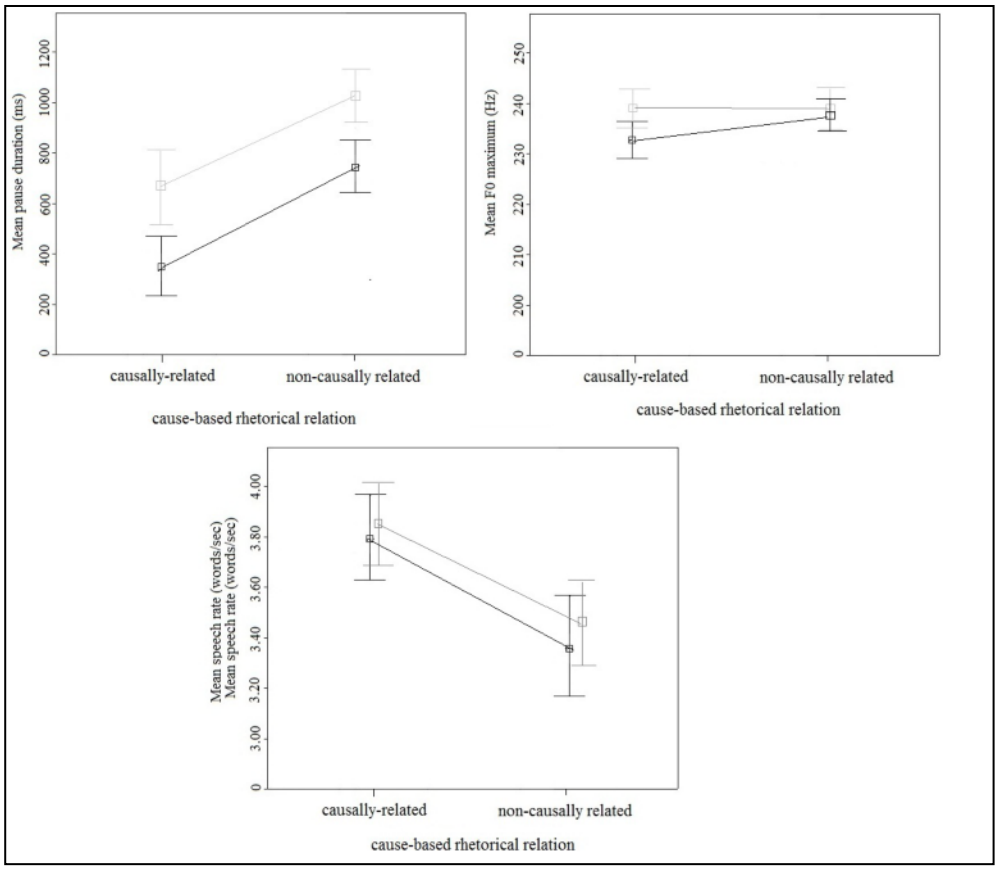

Figure 4. *Mean and Standard Error Values of Pause Duration (top left), F0-maximum (top right), and Speech Rate (bottom) as a Function of Cause-based Rhetorical Relation and Position of the Second Segment (initial: gray; non-initial: black) for all Speakers.*

Table 2. *Mean And Standard Error Values of Pause Duration (Ms), F0-Maximum (Hz), and Speech Rate as a Function of Source Relation (Semantic Vs. Pragmatic) and Position of the Second Segment of the Pair*

		sematic	pragmatic
initial	pause duration (ms)	552	539
	F0-maximum (Hz)	234	236
	speech rate (words/sec)	3.37	3.32
non-initial	pause duration (ms)	374	361
	F0-maximum (Hz)	238	235
	speech rate (words/sec)	3.26	3.33

To summarize, the data in this section show that causally related sentences are articulated with shorter in-between time interval and faster speech rate than non-causally related sentences, suggesting that causal relations are connected more strongly than non-causal relations. The results replicate and extend earlier findings for English native speakers. As stated earlier, previous research has found that native speakers of English produce sentences that are causally related with shorter pause durations and at a faster rate than sentences that are non-causally related (den Ouden et al., 2009; Sanders & Noordman, 2000)

3.4.3. Effect of Nuclearity

In this section, we examined the prosodic realization of local text structure. In particular, we intended to explore the degree to which the nuclearity (or information status) of the segments could systematically affect values of the three measures of prosody; i.e., pause duration, F0 maximum and speech rate. Table 3 reports the prosodic characteristics of the nucleus and satellite segments in relation with the position of the second segment. We did not

include hierarchy as an independent factor because of too low frequencies in some cells of the matrix. It can be observed that none of the prosodic measures are affected by nuclearity: there are hardly any differences in pause duration, F0 maximum and speech rate between the nucleus and satellite segments. Results of two-way ANOVAs indicated that the effect of nuclearity on pause duration ($F(1,68)=0.26$; $\varrho=0.61$), F0 maximum ($F(1,68)=1.76$; $\varrho=0.18$) and speech rate ($F(1,68)= 2.39$; $\varrho= 0.12$) was not significant. Position affected pause duration ($F(1,68)= 49.53$; $\varrho< 0.001$), but not F0 maximum ($F(1,68)= 1.7$; $\varrho= 0.19$) and speech rate ($F(1,68)= 1.23$; $\varrho= 0.27$).

Table 3. *Mean and Standard error Values of pause Duration (ms), F0-Maximum (Hz), and Speech Rate as a Function of Nuclearity (nucleus vs. satellite) and Position of the Second Segment*

		nucleus	satellite
initial	pause duration (ms)	489	466
	F0-maximum (Hz)	232	229
	speech rate (words/sec)	3.11	3.19
non-initial	pause duration (ms)	319	327
	F0-maximum (Hz)	239	233
	speech rate (words/sec)	3.49	3.56

Thus, our findings for nuclearity do not agree with previous results found for native speakers of English. As explained earlier, it has been suggested that nuclear segments in native speakers' data are realized with a more prominent prosody, i.e., having longer preceding pauses, higher pitch range, and a slower articulation rate, than satellites (den Ouden et al., 2009; Sanders & Noordman, 2000; Schilperoord, 1996; Andreeva et al., 2014, 2016).

4. Discussion and Conclusion

The study reported here investigated the systematic use of prosodic cues to signal text structure and content in non-native spoken discourse data produced by Persian high proficient EFL learners. Previous examination of native speakers' discourse data suggests that speakers intentionally create a system of prosodic units to emphasize relationships between rhetorically related sections of the discourse and highlight global and local text structure. However, analysis of Persian EFL learners' data in the present study showed that speakers were unable to consistently manipulate prosodic characteristics of discourse in native-like manner to signal global and local text structure as well as rhetorical relations.

Although a precise comparison between native speakers' data and the Persian learners' data is impossible (due to differences in methodology, text materials, etc.), an overview reveals that while native English spoken discourse features a multiple cuing system operating concurrently at different levels of the discourse structure and used to accentuate subdivisions in the substantive content of the text, Persian learners' spoken data lack a structured multi-level prosodic system to cue text structure and content. In particular, the results of the present investigation showed that the intonational cue (pitch variation) to prosodic prominence in discourse was completely absent in Persian speakers' spoken data. Speakers showed no tendency to use variation in pith range as a structured prosodic cue to signal the global structure, rhetorical relations or the local structure of English text. The lack of pitch change structuring in Persian speakers' data has been previously reported for other L2 learners. Earlier studies of Mennen (1998), Lochwyn (1999) and Wessels & Lawrence (1992), have all found that overall reduction in pitch range is a critical feature of non-native spoken discourse. Furthermore, these studies suggest that the lack of (or

significantly narrower) pitch range variation in L2 speakers may be related to a lack of confidence in in the new language. As for the temporal cues to prosodic representation of discourse, i.e., pause duration and speech rate, no simple comparison can be made between the non-native Persian speakers' data of the present study and data from English native speakers. For example, no features that have consistently appeared in the native speakers' spoken discourse were entirely lacking in the Persian learner group. Speakers made a consistent use of variation in pause duration to signal global text structure. In addition, they varied pause duration as well as speech rate consistently in native-like manner to differentiate between causally related vs. non-causally related segments. However, unlike native speakers, Persian EFL learners were unable to consistently manipulate speech rate to signal global text structure and pause duration and speech rate to signal local text structure.

Overall, the findings of the present study may be interpreted as suggesting that there is little evidence of a structured prosodic system in the production of Persian EFL learners to signal English text structure and content.

With respect to the pedagogical implications of the current study, as stated by Davis (2004), we assume that the ability to successfully utilize prosodic features for the representation of discourse organization and content is crucial for the development of interactional competence and has broad practical implications for second language learning and teaching. Incorrect use of prosodic features in reading a text will not only cause communication errors but also intensify misunderstandings in text interpretation. However, as Wrembel (2007) notes, despite a consensus regarding the significance of prosodic features for successful communication "prosody still appears to be the "problem child" from the pedagogical perspective" (p. 189). Certainly, a cursory review of EFL textbooks shows limited if any discussion of the role of

prosodic features in in signaling text structure and content. Jilka (2007) suggests that we might teach learners "conscious control" of features such as pitch range and also suggests the use of speech technology to facilitate this. In addition, as noted by Pickering (2004), while sentence-level prosody is related to universal cognitive constraints, the organization of longer structural units may be a learned skill and subject to cross-linguistic variation. Thus, construction of discourse prosody could be us8fully included in training curricula. Specifically, Persian learners' inability to use pitch range variation in their spoken discourse can be directly addressed through the use of standard exercises from theater training designed to increase confidence and encourage speakers to explore their voice range. Vocal warm-ups such as "call and response" exercises also help students to produce an extended pitch range as they practice calling a person from a distance (Lochwyn, 1999; Wessels & Lawrence, 1992).

The sample for the present study comprised of two newspaper texts. Therefore, research studies with larger sample size would be required to ensure appropriate generalization of the findings of the study.

References

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning*, *42*, 529–555.

Anderson-Hsieh, J., & Venkatagiri, H. (1995). Syllable duration and pausing in the speech of Chinese ESL speakers. *TESOL Quarterly*, *42*, 807–812.

Andreeva, B., Barry, W., & Koreman, J. (2014). A Cross-Language Corpus for Studying the Phonetics and Phonology of Prominence. Reykjavik, *The 9th Edition of the Language Resources and Evaluation Conference*, 26-31.

Andreeva, B., Barry, W., & Koreman, J. (2016). Local and Global Cues in the Prosodic Realization of Broad and Narrow Focus in Bulgarian. *Phonetica*, 73, 256-278.

Asher, N., & Lascarides, A. (2003). *Logics of Conversation*. Cambridge, UK: Cambridge UPress.

Bateman, J. A., & Rondhuis, K. J. (1997). Coherence relations: Towards a general specification. *Discourse Processes, 24*, 3–49.

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer (Version 4.3.01) [Computer program]. http://www.praat.org/.

Braun, B. (2015). What causes the activation of contrastive alternatives, the size of focus domain or pitch accent type? *Proceedings of the 18th ICPhS*, Glasgow, UK.

Calhoun, S. (2010). How does informativeness affect prosodic prominence?. *Lang Cogn Process, 25*, 1099-1140.

Davis, C. (2004). Developing awareness of crosscultural pragmatics: The case of American/German sociable interaction. *Multilingua,* 23, 207-231.

den Ouden, H. (2004). Prosodic Realizations of Text Structure (PhD Dissertation). University of Tilburg.

den Ouden, H., & Terken, J. (2001). Measuring pitch range. In Proc. 7th Eur. Confer. Speech Commun. Technol., Aalborg, Danmark, 91–94.

den Ouden, H., Noordman, L., & Terken, J. (2009). Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports. *Speech Communication*, 51, 116–129.

Hirschberg, J., & Grosz, B. (1992). Intonational features of local and global discourse structure. *Proceedings Speech and Natural Language Workshop*. Harriman, New York, pp. 441–446.

Hirschberg, J., & Nakatani, C. (1996). A prosodic analysis of discourse segments in direction-giving monologues. 3*4th Proceedings of the Annual Meeting Association for Computational Linguistics*, Santa Cruz, 286–293.

Jilka, M. (2007). Different manifestations and perceptions of foreign accent in intonation. In J. Trouvain & U. Gut (Eds). *Non-native prosody: Phonetic description & Teaching Practice.* (pp.77-96). Berlin: Mouton de Gruyter.

Lehiste, I. (1975). The phonetic structure of paragraphs. In Cohen, A., Nooteboom, S. (Eds.), *Structure and Process in Speech Perception* (pp. 195–203). Springer, Berlin.

Lochwyn, T. (1999). Theater exercises in the ESL classroom. Unpublished paper.

Mann, B., & Thompson, S. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text,* 8, 243–281.

Mennen, I. (1998). Can language learners ever acquire the intonation of a second language?. *Proceedings of the ESCA workshop on speech technology in language learning* (pp. 17–20). Marholmen, Sweden: International Speech Communication Association.

Noordman, L., Dassen, I., Swerts, M., Terken, J. (1999). Prosodic markers of text structure. In van Hoek, K., Kibrik, A., Noordman, L. (Eds.), *Discourse Studies in Cognitive Linguistics* (pp. 133–145). Benjamins, Amsterdam,

Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly*, 35, 233–255.

Pickering, L. (2004). The structure and function of intonational paragraphs in native and non-native speaker instructional discourse. *English for Specific Purposes*, 23, 19–43.

Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics* (pp. 223–255). Academis Press, New York,

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, 14, 423–441.

Rounds, P. (1987). Characterizing successful classroom discourse for NNS teaching assistant training. *TESOL Quarterly*, 21, 643–672.

Sanders, T., Noordman, L. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes* 29, 37–60.

Sanders, T., Spooren, W., Noordman, L. (1992). Toward a taxonomy of coherence relations. *Discourse Processes,* 15, 1–35.

Schilperoord, J. (1996). It's About Time. Dissertation, Utrecht University.

Silverman, K. E. A. (1987). *The structure and processing of fundamental frequency contours* (Doctoral dissertation). Cambridge, UK: Cambridge University.

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *J. Acoust. Soc. Am.* 101, 514–521.

Sweetser, E. (1990). *From Etymology to Pragmatics*. Cambridge University Press, Cambridge.

Thorsen, G., & N. Gronnum, (1985). Intonation and text in standard Danish. *J. Acoust. Soc. Am.* 80, 1205–1216.

Tyler, A., & Davies, C. (1990). Cross-linguistic communication missteps. *Text*, 10, 385–411.

Tyler, A., Jefferies, A., & Davies, C. (1988). The effect of discourse structuring devices on listener perceptions of coherence in nonnative university teachers' spoken discourse. *World Englishes*, 7, 101–110.

van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language & Speech*, 38, 253–265.

van Donzel, M. (1999). Prosodic Aspects of Information Structure in Discourse. PhD dissertation, University of Amsterdam.

Wennerstrom, A. (1994). Intonational meaning in English discourse. *Applied Linguistics*, 15, 399–421.

Wessels, C., & Lawrence, K. (1992). Using drama voice techniques in the teaching of pronunciation. In A. Brown (Ed.), *Approaches to pronunciation teaching* (pp. 29–37). London: McMillan.

Wrembel, M. (2007). Metacompetence-based approach to the teaching of L2 prosody: Practical implications. In J. Trouvain & U. Gut (Eds). *Non-native prosody: Phonetic description & Teaching Practice.* (pp. 189-210). Berlin: Mouton de Gruyter.

Yule, G. (1980). Speakers' topics and major paratones. *Lingua,* 52, 33–47.