INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Acoustic cues to topic and narrow focus in Egyptian Arabic

Dina El Zarka[1], Barbara Schuppler[2], Francesco Cangemi[3]

[1]Department of Linguistics, University of Graz, Austria
[2]Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria
[3]IfL Phonetics, University of Cologne, Germany

dina.elzarka@uni-graz.at, b.schuppler@tugraz.at, fcangemi@uni-koeln.de

Abstract

This study investigates acoustic cues (duration, scaling and alignment of peaks and valleys) to the prosodic realization of *topics* and *narrow subject foci* in a declarative SVO sentence in Egyptian Arabic. Morpho-syntactically identical sentences were elicited in appropriately designed contexts from 18 native speakers by means of a question-answer paradigm. The results show that the stressed syllable of a focused word is longer than the stressed syllable of the same word in topic condition. Additionally, the peaks of foci are generally scaled higher than those of topics. These differences clearly point to varying degrees of prosodic prominence. Furthermore, the alignment of the F0 peak and the subsequent low endpoint of a rising-falling tonal contour is earlier in foci than in topics, indicating that focus is signaled by an early sharp fall whereas the falling part of the tonal gesture starts later and is shallower in the case of a topic. Overall, our results suggest that narrow subject foci and topics tend to be associated with different pitch events.

Index Terms: Acoustic Features, Prosody, Egyptian Arabic, Narrow Focus, Topic

1. Introduction

An uncontroversial fact about Egyptian Arabic (EA) prosody is that every content word is accented [1, 2, 3, 4]. Within a standard autosegmental-metrical model, the intonation of EA has been described with one type of pitch accent [1], a rising gesture with a stably aligned low target at the beginning of the stressed syllable and a high target whose alignment is dependent on the structure of the stressed syllable. According to [2, 5], topics are characterized by rising or rising-level contours whereas foci are associated with (rise-)falls, implying that the alignment of the tones not only varies with syllable structure but is subject to functional variation depending on the information structure of the word carrying the pitch accent. Whereas the proposal in [2] was predominantly based on qualitative analysis, the present study aims at providing quantitative evidence for potential categorical differences between pitch events associated with topics and foci.

So far, the acoustic features of topics in EA have not been investigated in detail. Other experimental studies have only studied the acoustic characteristics of accents associated with target words uttered in different focus conditions (broad focus, narrow information focus, narrow contrastive focus). All studies agree that focus may enhance the excursion size of an accent [1, 2, 3, 4, 6, 7].

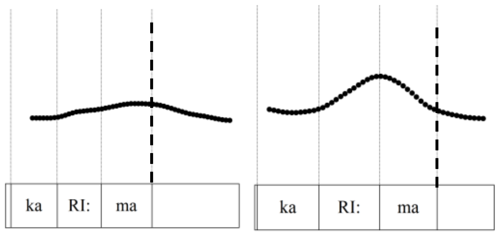

Figure 1. *F0 tracings of a rising accent on a topic constituent (left panel) and a rise-fall on a segmentally identical focus constituent (right panel). Dashed lines indicate the boundary of the word.*

Concerning the alignment properties of the accent, they have yielded different results: whereas [1, 8] found no alignment difference under focus, [6] found that some speakers use alignment to differentiate between narrow and broad focus while others do not. According to the pilot experiment described in [2], foci also exhibit longer overall durations than topics in EA.

Figure 1 shows examples from our data set for rising-falling tonal contours associated with a non-contrastive topic (left panel) and a non-contrastive focus (right panel). Some observations can be made based on the figure: (i) the accent peak in a focus constituent is aligned earlier than in a topic accent. (ii) The duration of the stressed syllable is longer in focus than in topic condition. (iii) Finally, the excursion size of the accent is larger in focus than in topic condition. The present study aims at investigating how these acoustic cues contribute to the overall shape differences observed for focus vs. topic based on a production study with a large number of speakers. More specifically, we investigate the effect of information structure (i.e., of sentence-initial *narrow (subject) focus* vs. sentence-initial *topic*) on syllable duration, the alignment of the rise endpoint (H) and of the fall endpoint (L), and H scaling. Thereby, we extend the method used in [6] (who studied broad vs. narrow focus) to the investigation of topics and additionally analyze durational features. This will allow us to test the following hypotheses:

http://dx.doi.org/10.21437/Interspeech.2019-1189

- H1: Focused words exhibit longer durations than their topic counterparts; the durational effect is expected to arise predominantly, but not exclusively, from the lengthening of the stressed syllable.
- H2: The peak (H) of a focus accent is scaled higher than the H of a topic accent.
- H3: Whereas H is aligned around the end of the stressed syllable in a focus constituent, it is aligned later, towards the end of the target word, in topic constituents.
- H4: Whereas the low turning point (L) following the peak is aligned near the end of the target word in a focus constituent, it is aligned further away from the target word boundary in a topic constituent.

2. Method

2.1. Data and speakers

By means of a question-answer paradigm, target sentences were elicited from 18 native speakers of EA (11 female, 7 male) aged between 22 and 78 who were born and raised in Cairo or Alexandria. All but one had university-level education and learned English and/or French as a second language, but no one used any of these languages regularly.

Narrow focus and topic stimuli were elicited in a contrastive and a non-contrastive context by means of 6 mini-dialogues. As our hypotheses only pertain to the difference between *topics* and *foci*, irrespective of whether they are contrastive, we utilized the data from all elicited narrow focus and topic conditions in this study. However, to account for potential effects of contrast, we added the latter as a control factor in the statistical analysis (see §2.3). Participants were presented with the target sentence on a computer screen and listened to a pre-recorded question stimulus that was meant to trigger a specific information structure in the target answer.[1] Each target sentence consisted of three words: subject, verb and object, with the target word (TW), the subject, being a female proper name.

(1)

(a) **target word as focus**

A: *ħali:ma* *najjimit ama:ni.*. '**Halima** put Amani to bed.'

(non-contrastive)	(contrastive)
Q: *ama:ni fi sa:biʕ no:ma, mi:n najjimha nnahardˤa?*	Q: *mi:n fi:hum najjim Ama:ni? ħali:ma walla nabi:la?*
'Amani's fast asleep. Who put her to bed today?'	'Which one of them put Amani to bed, Halima or Nabila?

(b) **target word as topic**

A: *ħali:ma* *najjimit ama:ni.* '**Halima** put Amani to bed.'

(non-contrastive)	(contrastive)
Q: *ħali:ma ʕamalit ʔe:?*	Q: *nabi:la ħadˤdˤarit isˤ-sˤufra wi ħali:ma ʕamalit ʔe:?*
'What did Halima do?'	'Nabila set the table and what did Halima do?'

[1] This data was collected as part of a larger dataset, already employed in [6] to study speaker-specificity in the encoding of a different contrast.

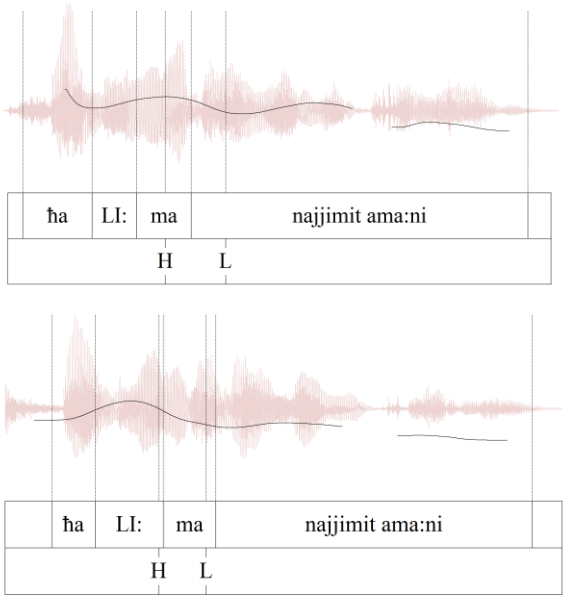

Figure 2: *Examples of annotated contours of a whole utterance with a topic (upper panel) and a narrow focus (lower panel). Dashed lines indicate word boundary.*

Example (1) shows morpho-syntactically identical answers to questions triggering a topic-comment partition (1a) or a narrow focus-background partition (1b). In total, we produced a set of 6 target sentences with a trisyllabic TW. The TW was stressed on the penultimate syllable and contained the long vowel /i:/ surrounded by either sonorants or /h/. Each target sentence was read three times in four conditions by 18 speakers (17*6*4*3 = 1224 plus 1*6*3*3 = 54),[2] yielding a total of 1278 recorded sentences. We excluded all items with poor recording quality, disfluencies and cases in which no clear rise was identifiable (n=31). One speaker produced cleft sentences under contrastive focus condition, which were excluded from analysis. In addition, we also excluded all items with a break after the TW (n=135)[3], as breaks have a major impact on duration and also potentially influence the behavior of tones. Finally, we excluded outliers (97% confidence interval) and ended up with 1029 TWs for the analysis.

2.2. Acoustic measurements

The TWs were segmented into syllables, F0 tracks were extracted automatically, manually corrected and smoothed using *mausmooth* [9] in *Praat* [10]. Figure 2 shows smoothed F0 tracks of two sentences given in Example 1 for topic (upper panel) and focus (lower panel), a phonological transcription and syllable segmentation of the TW (stressed syllable capitalized). The second tier shows the tone labels.

[2] Due to recording problems, one condition is missing for one speaker.

[3] Breaks occurred in all subcategories, but were more frequent in foci: contrastive focus (25%), narrow information focus (47%), contrastive topic (11%) and non-contrastive topic (17%). One speaker (F05) used breaks more frequently than others, but in her case, breaks were distributed almost evenly across information structure types. Since the present study focusses on pitch events, these aspects will not be further detailed in this paper (see also §2.2 and §4).

From the smoothed contours, the following turning points were detected using R [11]: For the high turning point H, maximum F0 was detected within a window starting at the beginning of the TW to 150ms after the end of the stressed syllable. For the low turning point L, the second derivative maximum was extracted, from H to 300ms after the end of the stressed syllable. Then the results of the automatic detection were manually verified and corrected where necessary. The elbows were used to characterize the F0 rises and falls in terms of scaling (in Hz) and alignment.

Based on prior work [1, 2], we assumed that the low turning point that started the rise was fairly stably aligned with the beginning of the stressed syllable. To answer the research question of this paper, we thus only needed to consider the variation of H and the following low turning point. H timing was calculated proportionally to the duration of the stressed syllable and L timing proportionally to the duration of the stress foot ranging from the start of the stressed syllable to the end of the TW. Values below 1 thus indicate alignment of H within the stressed syllable and of L within the TW. Values above 1 indicate alignment of H in the post-stress region and of L beyond the end of the TW.

As can be seen in Figure 2, a topic-comment sentence and a narrow focus-background sentence also exhibit overall tune differences. While pitch range is usually – but not necessarily – compressed after a narrow focus, the usual downdrift that is typical for all-new utterances [1, 2, 3] may be suspended or even replaced by upstepped accents in a topic-comment utterance [2, 12]. As our present paper investigates only the hypothesis that monomial topic and focus constituents are typically associated with different pitch events, we will not pursue the issue of whole tunes here.

2.3. Statistical analysis

In order to investigate whether the acoustic measurements are affected by information structure, we built six linear mixed effects regression models, one for each acoustic (dependent) variable: (*Initial*-, *Stressed*- and *Final*-) *Sylldur, H_scale, H_align, L_align*. In all models, information structure was represented by the independent variable TYPE, with the categories *focus* (F) and *topic* (T). This variable is assumed to predict that a focus would be associated with a rise-fall exhibiting a steeper falling gesture than a topic [2, 5]. It has, however, been shown for other languages that contrast enhances prominence-lending acoustic cues such as duration and F0 height [e.g., 13, 14]. Likewise, contrastive topics may be prosodically distinguished from non-contrastive ones in similar ways [e.g., 15]. Similar effects can also be expected for EA. We therefore added CONTRAST as a control variable to our models. Finally, each model included the independent variable REPETITION (3 values) and the random variables SENTENCE (6 values) and SPEAKER (18 values). The random variables are both incorporated in terms of random intercept models. A detailed analysis of speaker-specific strategies, as for instance presented in [6], is left for future work.

For building the linear mixed effects regression models, we used the *lmer function* of the *lme4 package* in *R* [11]. We started by including all independent variables and their interactions into the models and then successively excluded non-significant factors and interactions as long as the model would still improve, given their AIC values and degrees of freedom [16, 17].

3. Results and discussion

3.1. Duration

Table 1: *Final model for syllable duration*

	est	**s.d.**	**t**	**p**
Intercept	155.18	7.21	21.51	<0.001
Type(T)	-18.71	1.23	-15.18	<0.001
Contr(yes)	-0.57	1.22	-0.47	0.64

Our first hypothesis (H1) was that foci and topics differ in duration. In all statistical models concerning durational measures, the random variables SPEAKER and SENTENCE were significant. In the statistical models for *Initial*- and *Final-Sylldur*, TYPE did not show significant effects. Instead, we found that initial syllables tend to be significantly shorter for contrastive TWs (est = -10.41, t = -6.97; p<0.001) and that final syllables tend to be significantly longer the more often speakers have repeated the target sentence (est = 3.09; t = 2.59; p<0.01). By contrast, *Stressed-Sylldur* (cf. full model shown in Table 1), resulted to be significantly affected by TYPE with no other independent variable having a significant effect.

The fact that the duration of the final syllable was not affected by TYPE is especially relevant as it indicates that there is no phrasing difference between topics and foci in the analyzed data set. Compared to a pilot experiment [2], which only reported a significant effect of TYPE on overall word duration, the results from the larger data set used in the present study permit a more detailed interpretation, suggesting that it is the stressed syllable that is responsible for the durational differences.

3.2. Peak height

Table 2: *Final model for H-scaling*

	est	**s.d.**	**t**	**p**
Intercept	230.64	16.54	13.94	<0.001
Type(T)	-27.61	1.80	-15.30	<0.001
Contr(yes)	-2.93	1.76	-1.66	0.1
Rep(2)	0.78	1.50	0.52	0.6
Rep(3)	3.86	1.51	2.56	<0.05
Type(T):Contr	10.93	2.48	4.41	<0.001

Our second hypothesis (H2) was that the peak of a focus accent is scaled higher than the peak of a topic accent. In order to test H2, we built a statistical model with H_scale as independent variable (cf. Table 2). TYPE resulted to be highly significant: peaks of focus accents tend to be significantly higher (μ=229Hz, est = -27,61; t = -15.30; p<0.001) than peaks of topic accents (μ=210Hz). CONTRAST on its own showed no significant effect, but its interaction with TYPE did. For non-contrastive TWs the difference in peak height between focus (μ=230Hz) and topic (μ=207Hz) is significantly larger (est = 10.93; t = 4.41; p<0.001) than for contrastive TWs (μ(F)=226Hz, μ(T)=212Hz). For additional analysis, we ran our models on two subsets of the data, one with focused TWs (N= 499) and the other with topic TWs (T=519) only. In both subsets, CONTRAST remains a significant variable, however, its effect is smaller in the focus subset (est = -4.92, t = -3.11, p<0.01) than in the topic subset (est = 7.93, t = 4.48, p<0.001). In general, these results support

H2, but they also show that contrast as an independent variable also raises the peak height in a topic whereas it slightly lowers peak height in a focused TW. This effect deserves further investigation in future work.

3.3. Alignment of H and L

Table 3: *Final model for H-alignment*

	est	s.d.	t	p
Intercept	0.96	0.07	13.94	<0.001
Type(T)	0.32	0.01	25.95	<0.001
Contr(yes)	-0.00	0.01	-0.30	0.77

Our third hypothesis (H3) was that whereas the F0 of the peak is aligned around the end of the stressed syllable in a focus constituent, it is aligned towards the end of the target word in a topic constituent. In order to test H3, we built a model with *H_align* as dependent variable (cf. Table 3). The statistical results validate H3 as we found that TYPE had a significant effect on *H_align*: While the mean peak alignment tends to be within the stressed syllable in a focus accent (μ=0,92), it tends to be outside the syllable towards the end of the TW in a topic accent (μ=1,25, est = 0.32; t = 25.95; $p<0.001$), suggesting that the rise covers most of the stress foot.

Table 4: *Final model for L-alignment*

	est	s.d.	t	p
Intercept	1.17	0.03	37.94	<0.001
Type(T)	0.28	0.02	16.09	<0.001
Contr(yes)	-0.04	0.01	-2.17	<0.05
Rep(2)	-0.02	0.01	-1.77	0.08
Rep(3)	-0.03	0.01	-1.77	0.09
Type(T):Contr	0.04	0.02	1.82	0.07

Our fourth hypothesis (H4) was that whereas the low target following the peak is aligned near the end of the target word in a focus constituent, it is aligned further away from the target word boundary in a topic constituent. In order to test H4, we built a model with L_align as dependent variable (cf. Table 4). In this model TYPE and CONTRAST resulted to be the significant variables. As predicted, L lies in the vicinity of word end in the case of foci (μ=1,12) and beyond in the case of topics (μ=1,43, est = 0.28; t = 16.09; $p<0.001$).

The comparison of the results from H- and L-alignment shows that whereas contrast is not a significant factor for peak alignment, it significantly affects the alignment of the L (est = 0.04; t = -2.17; $p<0.05$). This effect, however, is smaller than the effect of information structure (est = 0.28; t = 16.09; $p<0.001$) as illustrated in Figure 3. We thus conclude that information structure is the main factor for the alignment of both turning points.

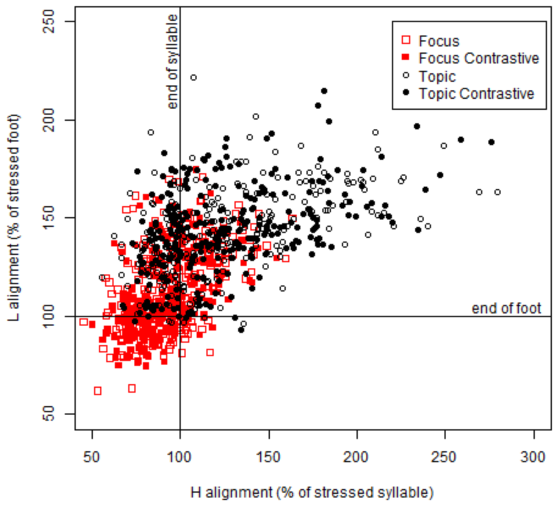

Figure 3: *Scatterplot for H- and L-alignment for focus (red squares) and topic (black circles), for contrastive contexts (filled symbols) and non-contrastive contexts (empty symbols).*

4. General discussion and conclusion

In this paper, we investigated how different acoustic cues (i.e., syllable duration, H scaling, H and L alignment) are affected by information structure in Egyptian Arabic. We analyzed the TWs of 1029 sentences produced by 18 speakers in a question-answer paradigm. All four hypotheses put forward in Section 1.2 could be validated: Foci tend to have longer stressed syllables and higher peaks than topics. The peaks as well as the subsequent low turning points are aligned earlier in foci than in topics.

In prior work [2], focus was shown to enhance word duration. The results of the present study indicate that this effect is due to the longer duration of the stressed syllable. As far as peak scaling is concerned, our results confirmed the findings of prior studies on focus in EA with respect to the effect of narrow focus [7, 1, 8, 2, 6], however, across a larger number of speakers and tokens. [6] had already shown that for some speakers, peaks are higher in narrow focus than in broad focus (all-new). According to the results of the present study, this effect is even stronger when narrow foci are compared to topics. Additionally, they suggest that contrast may override this tendency as peaks in contrastive topics were as high as in foci for some speakers. As duration and pitch height are prominence-related features, we conclude that focus and ‘syntagmatic’ contrast may be reflected in stronger prominence. Regarding the alignment of the turning points, our results confirm the hypothesis that topic accents and focus accents differ in shape [2, 5], with focal accents exhibiting a steeper fall, and that shape is predominantly determined by information structure rather than explicit contrast.

In future work, we will also investigate the issue of tunes as associated with the sentences as a whole, including the question of post-focal pitch range compression and post-topic upstep, and study the impact of contrast on topics and foci in some detail. We expect to find different speaker-specific strategies as suggested by visual inspection of the data similar to what [6] found for *narrow* and *broad focus.* The conclusion we can already draw from the acoustic analysis presented here is that the accent shapes of *topics* and *narrow foci* potentially belong to different pitch events.

5. Acknowledgements

The work by Barbara Schuppler was supported by the Elise Richter grant (V638 N33) from the Austrian Science Fund. Work by Francesco Cangemi was supported by the German Research Foundation through the Collaborative Research Center “Prominence in Language” (SFB 1252). We would like to thank Harriet Hanekamp for her assistance in post-processing the data.

6. References

[1] S. Hellmuth, *Intonational pitch accent distribution in Egyptian Arabic*. Dissertation. School of Oriental and African Studies: London, 2006.

[2] D. El Zarka, *On the interaction of information structure and prosody. The case of Egyptian Arabic*. Habilitation thesis. University of Graz: Graz, 2013.

[3] Kh. Rifaat, "The Structure of Arabic Intonation: A preliminary investigation", M.T. Alhawary and E. Benmamoun (eds), *Perspectives on Arabic Linguistics* XVII-XVIII, pp. 49–67, 2005.

[4] D. Rastegar-El Zarka, *Prosodische Phonologie des Arabischen*. Ph.D thesis, University of Graz: Graz, 1997.

[5] D. El Zarka, "Leading, Linking, and Closing Tones and Tunes in Egyptian Arabic: What a Simple Intonation System Tells Us about the Nature of Intonation", E. Broselow and H. Ouali (eds), *Perspectives on Arabic Linguistics* XXII-XXIII, pp. 57–73, 2011.

[6] F. Cangemi, D. El Zarka, S. Wehrle, Simon, S. Baumann and M. Grice, "Speaker-specific intonational marking of narrow focus in Egyptian Arabic", *Proceedings of Speech Prosody 2016,* pp. 335–339.

[7] K. Norlin, "A preliminary description of Cairo Arabic intonation of statements and questions", *Quarterly Progress and Status Report (QPSR)*, 1989, pp. 47–49.

[8] S. Hellmuth, "The (absence of) prosodic reflexes of the given/new distinction in Egyptian Arabic", J. Owens and A. Elgibali (eds), *Information Structure in Spoken Arabic*. London: Routledge, 2010, pp.165–188.

[9] F. Cangemi, *mausmooth* [Online], 2015. Available: http://phonetik.phil-fak.uni-koeln.de/fcangemi.html.

[10] P. Boersma and D. Weenink, Praat: doing phonetics by computer (version 6.0.37) [Online], 2018. Available: http://www.praat.org/.

[11] R Core Team R: a language and environment for statistical computing [Online]. R Foundation for Statistical Computing, Vienna, Austria, 2015. Available: https://www.R-project.org.

[12] D. El Zarka, B. Schuppler, "On the interplay of pragmatic and formal factors in the prosodic realization of themes in Egyptian Arabic" *Grazer Linguistische Studien (Graz Linguistic Studies)* 90, 2, pp. 33–106. (DOI:10.25364/04.45:2018.90.2).

[13] S. Baumann, J. Becker, M. Grice, D. Mücke, "Tonal and articulatory marking of focus in German", *Proceedings of the 16th International Conference of the Phonetic Sciences (ICPhS)*, Saarbrücken, Germany, 2007, pp. 1029–1032.

[14] J. Hanssen, J. Peters, C. Gussenhoven, "Prosodic effects of focus in Dutch declaratives", *Proceedings of Speech Prosody* 2008, pp. 609–612.

[15] B. Braun, B. Ladd, "Prosodic correlates of contrastive and non-contrastive themes in German", *Proceedings of the 8th European Conference on Speech Communication and Technology* 2003, pp. 789-792.

[16] R. H. Baayen, *Analyzing linguistic data. A practical introduction to statistics using R.* Cambridge: Cambridge Univ. Press, 2008.

[17] K. Johnson, *Quantitative Methods in Linguistics*. Oxford, UK: Blackwell Publishing, 2008.