# A CROSSLINGUISTIC STUDY OF PROSODIC FOCUS

*Yong-cheol Lee[1], Bei Wang[2], Sisi Chen[2], Martine Adda-Decker[3], Angélique Amelot[3], Satoshi Nambu[4], and Mark Liberman[1]*

[1]University of Pennsylvania, [2]Minzu University of China, [3]Sorbonne Nouvelle, [4]NINJAL

## ABSTRACT

We examined the production and perception of (contrastive) prosodic focus, using a paradigm based on digit strings, in which the same material and discourse contexts can be used in different languages. We found a striking difference between languages like English and Mandarin Chinese, where prosodic focus is clearly marked in production and accurately recognized in perception, and languages like Korean, where prosodic focus is neither clearly marked in production nor accurately recognized in perception. We also present comparable production data for Suzhou Wu, Japanese, and French.

***Index Terms*—** Focus, prosody, typology

## 1. INTRODUCTION

It has often been observed that different languages use different methods to different degrees in implementing different types of information-structural "focus" [1–5]. There are syntactic options such as cleft structures and scrambling of word order [6–10]; there may be relevant morphological markers [11–13]; and there are prosodic changes in phrasing or in the phonetic implementation of the focused element and the surrounding material [14–24].

It is well known that particular syntactic or morphological markers of information structure may be obligatory in some languages, but optional or non-existent in other languages. It is less commonly recognized that purely prosodic marking of focus may be much weaker in some languages than in others, to the extent that purely prosodic focus may be nearly absent as a general mechanism for communication of information structure.

Reliable quantitative cross-linguistic comparisons of prosodic focus are difficult, given all of the many relevant ways that languages can differ. We have developed a method based on corrective focus in digit strings, which makes such comparisons easier. Our method has the advantage that syntactic and morphological revisions are ruled out, so that only prosodic modulation can be used to mark the focused item; the materials are pragmatically uniform, so that all of the positions in the string are equally susceptible to focusing; and (except for substituting the appropriate digit names) the same materials can be used in every language.

We have applied this method in production and perception experiments on three languages (American English, Mandarin Chinese, and the Seoul and South Kyungsang varieties of Korean). Production data is available for three additional languages (Tokyo Japanese, Suzhou Wu, and Standard French), with perception results to follow.

## 2. PRODUCTION

### 2.1. Speech materials

We used 10-digit number strings, read as connected individual digits grouped as (NNN)-(NNN)-(NNNN) in the style of American telephone numbers. For five of the six languages, we created sets of 100 10-digit sequences for each language, designed so that each digit occurs ten times in each sequential position, and each pair of digits occurs once spanning each adjacent pair of positions. Speakers read these digit strings in isolation as a background broad-focus condition, and in a Q&A dialogue below for contrastive focus, where someone asks for confirmation of a version in which one of the digits is incorrect, and the speaker answers with a string in which the wrong digit is given correctly.

Q: Is Mary's number 215-418-5623?
A: No, the number is 215-41<u>7</u>-5623.

Due to constraints on speakers' time, in Suzhou Wu the production material was limited to 10 10-digit sequences, with each digit occurring ten times in each position, and each string read in six focus conditions, broad focus and focus in the 1st, 3rd, 4th, 5th, and 10th positions.

### 2.2. Speakers

In each language, a set of native speakers participated in a production experiment. In English, there were 2 males and 3 females; in Mandarin, 2 males and three females; in Seoul Korean, 2 males and 3 females; in South Kyungsang Korean, 2 males and 3 females; in Suzhou Wu, 3 males and 3 females; in Japanese, 1 male and 2 females; and in French, 3 females.

### 2.3. Recording procedure

Recordings were conducted in quiet conditions, with the speaker seated in front of a computer monitor with a head-mounted microphone. Number strings were presented to speakers through PowerPoint slides.

The broad-focus condition was recorded first, followed by the contrastive-focus condition. During the contrastive-focus session, speakers listened to prompt questions through headphones and read target phone-number strings displayed on the screen as answers.

### 2.4. Acoustic measurements

Digit boundaries were manually labeled, and based on those boundaries, we calculated word duration in seconds, mean intensity in dB, and mean pitch in Hz (or f0 range for Mandarin contour tones) of each labeled interval. The mean f0 and f0 range were converted to semitones.

### 2.5. Analyses and results

The phone-number strings in the broad-focus condition were directly compared with the same sequences in the contrastive-focus condition by the aggregate measures of mean f0 (or f0 range), duration, and mean intensity.

As Figure 1 shows, there was a clear and consistent effect of focus in all the parameters for English and in Mandarin Chinese. Focused digits were clearly indicated by greater duration, f0, and intensity. These languages also showed clear post-focus compression [15, 21, 23], with reduced duration, f0, and intensity on the post-focus digits.

In comparison, Seoul Korean exhibited no clear focus effects on corrected digits – prosodic modulation by focus was weak, ambiguous, and unclear. The amount of modulation by focus for mean f0 was about or less than half the size of English and Mandarin Chinese. The durational cues to focus positions were very small – on average, about 8.5 msec extra duration on the focused digits. The ambiguity of focus modulation is illustrated by mean f0: when position 1 was focused, position 2 was even higher; when position 4 was focused, position 5 was even higher; when position 7 was focused, position 8 was even higher. Mean intensity effects of Seoul Korean by focus were not clear at all. It seems like that there was no clear indication of contrast position.

Similar to Seoul Korean, South Kyungsang Korean showed relatively weak and ambiguous modulation by focus. Corrected digits were not clearly indicated in every position by increased mean f0, duration and mean intensity. Increased mean f0 was found in a neighboring position to the corrected digit. For example, when position 4 was focused, position 5 was even higher; focus on 7 made 8 higher; focus on 9 made 10 higher.

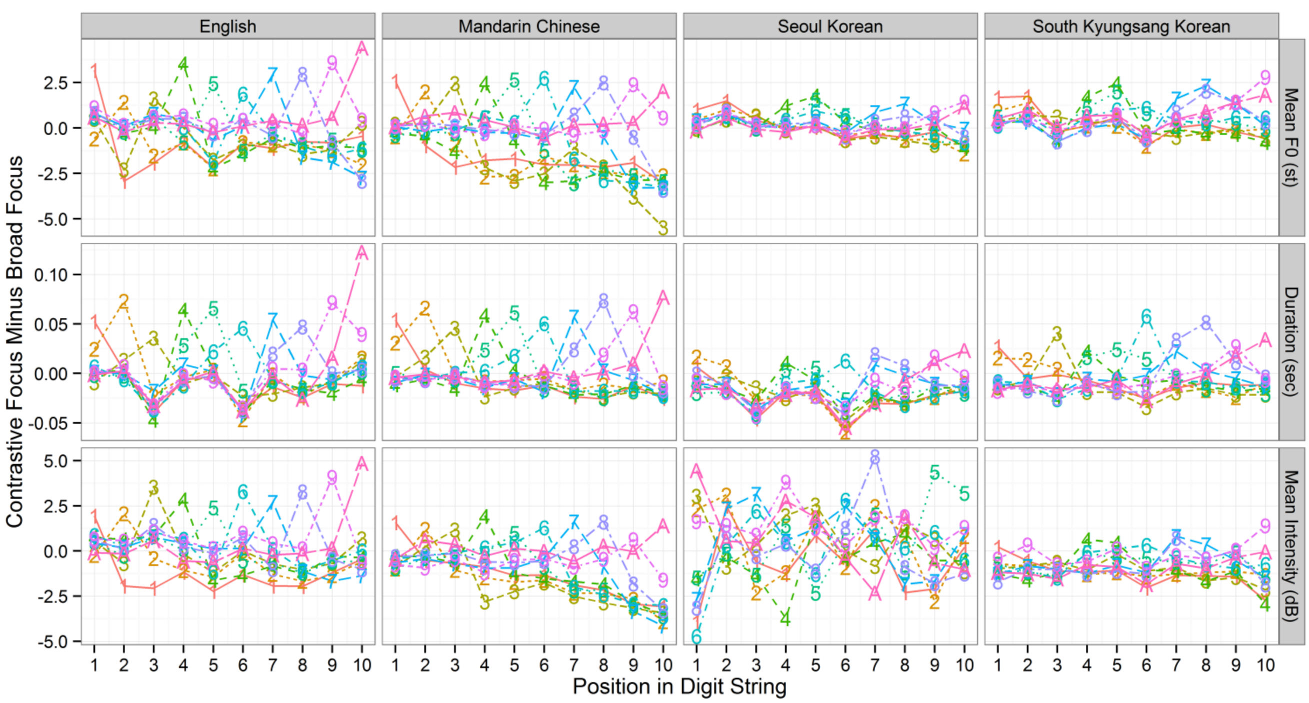


**Figure 1**. Prosodic differences between focused digits and broad-focus counterparts. Characters 1-A indicate focus position.

**Table 1**. Confusion matrix of contrastive focus perception in percent. Numbers highlighted in gray indicate correct identification rates. Dotted lines refer to a phrase in a phone-number string. (Top left panel: English, top right panel: Mandarin Chinese, bottom left panel: Seoul Korean, bottom right panel: South Kyungsang Korean)

| Target | Perceived | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 96 | 0 | 2 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 97 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 2 | 98 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 2 | 4 | 2 | 88 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 91 | 6 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 97 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 6 | 0 | 3 | 0 | 6 | 85 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 91 | 9 | 0 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 97 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 94 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 97 | 3 |
| 10 | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 96 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 94 |
| 1 | 49 | 13 | 17 | 4 | 2 | 10 | 4 | 0 | 1 | 1 | 72 | 5 | 8 | 0 | 0 | 5 | 3 | 5 | 0 | 3 |
| 2 | 11 | 27 | 15 | 2 | 5 | 1 | 20 | 0 | 10 | 8 | 0 | 21 | 36 | 8 | 5 | 10 | 3 | 0 | 13 | 5 |
| 3 | 11 | 4 | 69 | 2 | 4 | 7 | 4 | 0 | 0 | 0 | 3 | 5 | 56 | 0 | 0 | 3 | 0 | 5 | 28 | 0 |
| 4 | 10 | 11 | 6 | 49 | 5 | 10 | 5 | 4 | 1 | 1 | 3 | 0 | 3 | 77 | 13 | 0 | 3 | 0 | 0 | 3 |
| 5 | 1 | 0 | 5 | 6 | 15 | 32 | 27 | 6 | 6 | 1 | 0 | 5 | 3 | 5 | 56 | 15 | 0 | 5 | 10 | 0 |
| 6 | 0 | 0 | 4 | 20 | 13 | 55 | 5 | 0 | 2 | 1 | 5 | 3 | 5 | 26 | 3 | 54 | 3 | 0 | 0 | 3 |
| 7 | 11 | 4 | 1 | 12 | 0 | 4 | 54 | 11 | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 74 | 18 | 8 | 0 |
| 8 | 12 | 14 | 6 | 4 | 1 | 2 | 19 | 38 | 0 | 4 | 0 | 0 | 0 | 0 | 3 | 0 | 18 | 69 | 3 | 8 |
| 9 | 5 | 2 | 4 | 4 | 0 | 2 | 8 | 7 | 61 | 7 | 0 | 3 | 10 | 3 | 5 | 0 | 8 | 28 | 33 | 10 |
| 10 | 0 | 6 | 13 | 11 | 6 | 11 | 6 | 0 | 18 | 30 | 0 | 3 | 0 | 3 | 0 | 3 | 23 | 3 | 23 | 44 |

## 3. PERCEPTION

Given these striking differences in production, we were not surprised to see equally striking differences in perception among these three languages.

### 3.1. Stimuli, subjects, and listening procedure

A pilot perception experiment was conducted to examine the extent to which corrected digits are identified in each language. Thirty phone-number strings (three digits in each position) were used and randomized for each language. Listeners (English: 18, Mandarin: 11, Seoul Korean: 28, South Kyungsang Korean: 13) heard only the phrase with the correction, and were asked to identify which digit was corrected. This experiment was done online through Qualtrics.

### 3.2. Results

Table 1 above shows a confusion matrix for the identification of corrected digits in each language. Listeners identified the corrected digit 97.3% of the time for English and 94.9% for Mandarin Chinese. In contrast, corrected digits were identified at a rate of 44.6% for Seoul Korean and 55.6% for South Kyungsang Korean.

In both Seoul Korean and South Kyungsang Korean, false answers often named another digits within the same phrase, usually immediately preceding or following the correct answer. But even if we score by phrase rather than by digit, the overall identification rates are still only 68.7% for Seoul Korean and 82.3% for South Kyungsang Korean, as illustrated in Table 2.

**Table 2**. The phrase-by-phrase confusion matrix for the two varieties of Korean (Top panel: Seoul Korean, Bottom panel: South Kyungsang Korean):

| Target | Perceived | | |
|---|---|---|---|
| | 1st phrase | 2nd phrase | 3rd phrase |
| 1st phrase | 72 | 12 | 16 |
| 2nd phrase | 12 | 68 | 20 |
| 3rd phrase | 19 | 14 | 67 |
| 1st phrase | 68 | 10 | 21 |
| 2nd phrase | 9 | 83 | 9 |
| 3rd phrase | 4 | 4 | 92 |

## 4. WU, JAPANESE, AND FRENCH

We have gathered production data for three additional languages/varieties: Suzhou Wu, Tokyo Japanese, and Standard French.

Suzhou Wu and Tokyo Japanese are quite similar to Korean, in that measures such a duration, f0, and intensity do not show patterns clearly identifying the corrected digit. Suzhou Wu also shows F0 rising in the positions adjacent to the focused digit, especially when the focus is in the middle of a prosodic phrase. French, on the other hand, shows the focused digit well marked by increases in intensity, duration, and (to a lesser extent) F0.

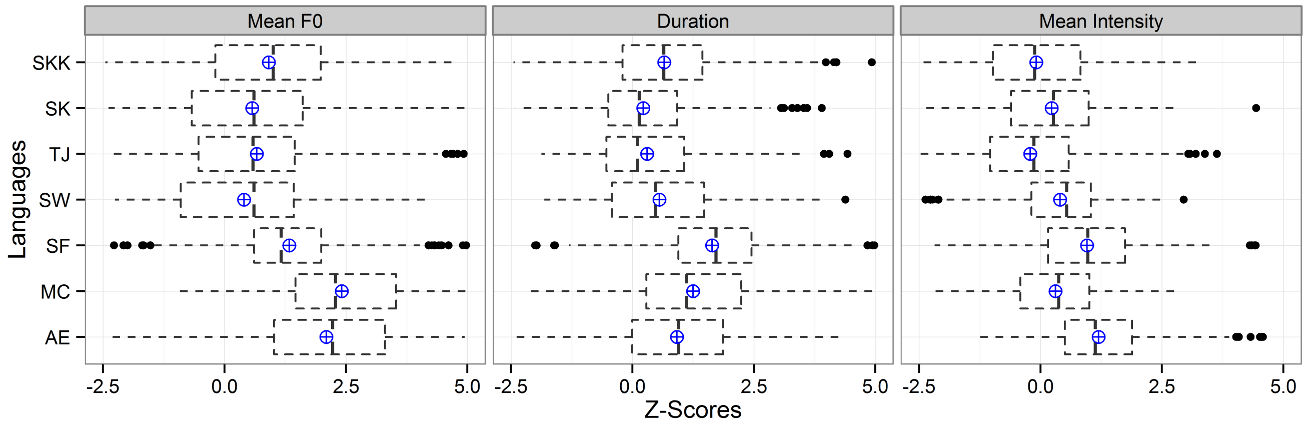


**Figure 2**. Distribution of z-scores for mean f0, duration, and mean intensity of focused digits (relative to mean and s.d. for unfocused instances of the same digit in the same position). (SKK: South Kyungsang Korean, SK: South Korean, TJ: Tokyo Japanese, SW: Suzhou Wu, SF: Standard French, MC: Mandarin Chinese, AE: American English)

We can see these patterns in Figure 2, which presents boxplots comparing focused to unfocused digits for all seven languages/varieties we have studied so far. Table 3 shows the median (z-score) values of focused digits:

**Table 3**. The median (z-score) values of focused digits**.**

| | SKK | SK | TJ | SW | SF | MC | AE |
|---|---|---|---|---|---|---|---|
| Duration | 0.64 | 0.13 | 0.10 | 0.48 | 1.73 | 1.19 | 0.95 |
| Intensity | –.26 | 0.24 | –.24 | 0.53 | 0.97 | 0.36 | 1.28 |
| Pitch | 1.00 | 0.62 | 0.60 | 0.61 | 1.17 | 3.13 | 2.96 |

Such numbers clearly do not tell the whole story – there are no doubt cues in local relationships of f0 and timing, for example. But in the four cases where we have perception results to compare, there is a good correspondence between perception accuracy and the degree of prosodic modification of focused digits relative to unfocused ones.

## 5. DISCUSSION AND FUTURE DIRECTIONS

The technique described in this paper allows a systematic and quantitative comparison of languages in terms of the prosodic marking of corrective or contrastive focus. We have shown that this method can be used for the study of perception as well as production, and that to a first approximation, the perception and production results are congruent.

The experiments done so far establish clearly that languages can differ greatly in how well their speakers communicate the location of corrective or contrastive focus by purely prosodic means. In particular, speakers of American English and Mandarin Chinese modulate duration, f0 and intensity in a clear way to signal the location of corrective focus, and listeners in those languages recognize the intended location with high accuracy (97.3% and 94.9%). In contrast, speakers of two varieties of Korean do not clearly mark corrective focus by prosodic changes, and listeners in turn have a much harder time correctly locating corrective focus from prosodic cues (44.6% and 55.6%).

In the near future, we will be adding perception results for the Suzhou Wu, Tokyo Japanese, and Standard French. The nature of this experimental paradigm will make it easy to add additional languages and varieties.

It will also be interesting to see whether native speakers of a language like Korean can accurately perceive the intended location of corrective focus in a language like Mandarin or English, and can themselves convey this information as an L2 speaker. If they are easily able to do so, this raises the question of why analogous kinds of modulation are not used to in speaking their own language. If their L2 production and perception abilities are limited in this respect, it raises questions for second-language instruction.

There is an important set of questions about the relationship of our results to variation on other typological dimensions, including traditional oppositions like “melodic” vs. “dynamic” stress as well as more modern ones like the presence or absence of “post-focus compression” [22].

Once we have a larger range of production and perception data, there will be an interesting opportunity to model the detailed relationship between prosodic modulation and perceptual accuracy. What features are listeners paying attention to? Can we use machine-learning techniques to match their successes (or failures)?

Finally, the term “focus” is commonly used to refer to many different things whose relationship remains a matter of discussion [26-28]. Perhaps this method can be part of a program to untangle these functions and signaling methods across languages and situations.

[Work partly supported by the Labex EFL (ANR.CGI)]

## 5. REFERENCES


[1] F. Ameka, “Focus constructions in Ewe and Akan: A comparative perspective”, in C. Collins and V. B. Manfredi, Eds., *Proceedings of the Kwa Comparative Syntax Workshop*, pp. 1–25, 1992.

[2] D. Büring, “Towards a typology of focus realization”, in M. Zimmermann and C. Féry, Eds., *Information Structure*, pp. 177–205, 2009.

[3] L.M. Hyman, “The interaction between focus and tone in Bantu”, in G. Rebuschi and L. Tuller, Eds., *The Grammar of Focus*, 1999.

[4] K. Lambrecht, *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*, Cambridge: Cambridge University Press, 1996.

[5] K. Nakanishi, “Prosody and information structure in Japanese: A case study of topic marker *wa*”, *Japanese/Korean Linguistics*, vol. 10, pp. 434–447, 2001.

[6] K, Balogh, “Focus and ‘only’ in Hungarian”, in B. D. ten Cate and H. W. Zeevat, Eds., *Logic, Language, and Computation Lecture Notes in Computer Science*, vol. 4363, pp. 31–44, 2007.

[7] M. Brody, “Some remarks on the focus field in Hungarian”, *UCL Working Papers in Linguistics*, vol. 2, pp. 201–225, 1990.

[8] E. Estebas-Vilaplana, *The Use and Realisation of Accentual Focus in Central Catalan with a Comparison to English*, Ph.D. dissertation, University College London, 2000.

[9] E. Vallduví, *The Informational Component*, PhD dissertation, University of Pennsylvania, 1990.

[10] E. Onea, “Exhaustiveness of Hungarian focus: Experimental evidence from Hungarian and German”, in A. Riester and E. Onea, Eds., *Focus at the Syntax-Semantics Interface*, pp. 53–68, 2009.

[11] I. Fiedler and S. Jannedy, “Prosody of focus marking in Ewe”, *Journal of African Languages and Linguistics*, vol. 34, pp. 1–46, 2013.

[12] M. Gordon, “The intonational realization of contrastive focus in Chickasaw”, in C. Lee, M. Gordon, and D. Büring, Eds., *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*, pp. 69–82, 2007.

[13] A. Rialland and S. Robert, “The intonational system of Wolof”, *Linguistics* vol. 39, pp. 893–940, 2001.

[14] S.-W. Chen, B. Wang, and Y. Xu, “Closely related languages, different ways of realizing focus”, *InterSpeech* pp. 1007–1010, 2009.

[15] W.E. Cooper, S.J. Eady, and P.R. Mueller, “Acoustical aspects of contrastive stress in question-answer contexts”, *Journal of the Acoustical Society of America*, vol. 77, pp. 2142–2156, 1985.

[16] L.J. Downing and B. Pompino-Marschall, “The focus prosody of Chichewa and the Stress-Focus constraint: a response to Samek-Lodovici (2005)”, *Natural Language and Linguistic Theory*, vol. 31, pp. 647–681, 2013.

[17] C. Féry, “Focus and phrasing in French”, *Audiatur Vox Sapientiae: A festschrift for Arnim von Stechow*, pp. 153–181, 2001.

[18] C. Ipek, “Phonetic realization of focus with no on-focus pitch range expansion in Turkish”, *ICPhS*, pp. 140–143, 2011.

[19] Y.-C. Lee and Y. Xu, “Phonetic realization of contrastive focus in Korean”, *Speech Prosody*, pp. 100033:1–4, 2010.

[20] Y. Xu, “Post-focus compression: Cross-linguistic distribution and historical origin”, *ICPhS*, pp. 152–155, 2011.

[21] Y. Xu, 1999. Effects of tone and focus on the formation and alignment of f0 contours. *Journal of phonetics*, vol. 27, pp. 55–105, 1999.

[22] Y. Xu, S.-W. Chen, and B. Wang, “Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family?”, *The Linguistic Review*, vol. 29, pp. 131–147, 2012.

[23] Y. Xu and C.X. “Phonetic realization of focus in English declarative intonation”, *Journal of Phonetics* vol. 33, pp. 159–197, 2005.

[24] B. Wang, L. Wang, and T. Qadir, “Prosodic realization of focus in six languages/dialects in China”, *ICPhS*, pp. 144–147, 2011.

[25] J. Hirschberg, “Communication and prosody: Functional aspects of prosody”, *Speech Communication* 36.1, pp. 31–43, 2002.

[26] B.A. Hockey, “The interpretation and realization of focus: An experimental study of English and Hungarian”, U. of Penn. PhD diss., 1998.

[27] I. Kruijff-Korbayova and M. Steedman, “Discourse and Information Structure”, *Journal of Logic, Language and Information*, vol. 12, pp. 249–259, 2003.

[28] J. Katz and E. Selkirk, “Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English”, *Language*, vol. 87, pp. 771–816, 2011.

[29] M.L G. Lecumberri and M.C. Abreu, “The manifestation of intonational focus in Castilian Spanish”, *Catalan Journal of Linguistics*, vol. 2, pp. 33-54, 2003.

[30] E.O. Selkirk, “Contrastive focus, givenness and the unmarked status of ‘discourse-new’”, *Acta Linguistica Hungarica* vol. 10, 55, pp. 331–346, 2008.

[31] A. Steube, “Correction by contrastive focus”, ZAS Papers in Linguistics 23, pp. 211–230, 2001.

[32] M. Rooth, “A theory of focus interpretation”, *Natural Language Semantics*, vol. 1, pp. 75–118, 1992.

[33] M. Breen et al., “Acoustic Correlates of Information Structure”, *Language and Cognitive Processes*, vol. 25, pp. 1044–1098, 2010.

[34] P. Touati, “Structures prosodique du suédois et du français”, *Working Paper 21*, Lund University, 1987.