STATIC AND DYNAMIC FEATURES OF ENGLISH MONOPHTHONGAL VOWELS PRODUCTION BY HIJAZI ARABIC L2 LEARNERS

Wael Almurashi,[1] Jalal Al-Tamimi,[2] and Ghada Khattab[3]

[1]Taibah University, Saudi Arabia; [2]Université Paris Cité, LLF, France; [3]Newcastle University, United Kingdom
wmurashi@taibahu.edu.sa; jalal.al-tamimi@u-paris.fr; ghada.khattab@newcastle.ac.uk

ABSTRACT

Dynamic cues in monophthongal vowels—in particular, vowel inherent spectral change (VISC)—have been found to add to our understanding of vowel identity and of crosslinguistic differences in vowel patterning. Less work has been carried out on the role they play in understanding second language vowel production. This study investigated the role of static versus dynamic cues (along with additional cues: vowel duration, F0, and F3) in the production patterns of English monophthongal vowels by Hijazi Arabic L2 learners (HA L2) compared with first language speakers of each of Hijazi (HA L1) and English (NE). Data were collected from three groups (20 HA L1, 20 HA L2, and 20 NE) producing target monophthong vowels in a word list with varied consonantal contexts. Results show that dynamic measures and additional cues provide insights into L2 production patterns that are not normally gleaned from static measures or F1 and F2 correlates alone.

Keywords: Static cues, dynamic cues, discriminant analysis, second language learning, Hijazi Arabic.

1. INTRODUCTION

Common practice in the investigation of second language (L2) vowel production relies on measuring the first two formants (F1 and F2) at the monophthong vowel's midpoint and making conclusions about the degree of second-language attainment. For some time, this static approach to vowel measurement was believed to provide an insight into the optimal acoustic characteristics of monophthong vowels [23]. However, many acoustic studies have since reported that measuring vowel formants from multiple locations (e.g., dynamic cues—in particular, vowel inherent spectral change [VISC] models) can provide more information and lead to better identification when using discriminant analysis, a statistical method that is used to predict listeners' vowel categorisation patterns [15]; [16]; [20]; [21]. The investigation of dynamic properties of vowel production remains understudied in the domain of L2 speech acquisition despite the fact that it constitutes a more sophisticated method for exploring the degree to which learners achieve target-like vowel production in their L2 [17].

VISC can be defined as the 'relatively slowly varying changes in formant frequencies associated with vowels, even in the absence of consonantal context' [21]. It is taken between two locations over the full duration of the vowel: one near the vowel's onset (at around 20%) and the other near the vowel's offset (at around 80%), to minimise the effects of surrounding consonants. VISC has been quantified through three approaches, namely: the offset model, which investigates the degree of spectral change; the slope model, which investigates the rate of spectral change; and finally, the direction model, which investigates the direction of spectral change [20].

In terms of offset, [20] found that speech dynamics are greater for speakers of languages that have a sparse monophthong vowel system (e.g., Chinese) than for those that have a dense vowel system (e.g., Korean and English) possibly due to the former having more variability in production [19]. Additionally, the production of English monophthongal vowels produced by Korean and Chinese L2 learners was found to be influenced by their L1 VISC (e.g., similar amount of VISC). The slope and direction models have also been found to provide a characterization of dynamic cues of monophthongal vowels and have been used to examine the intrinsic dynamic of monophthong vowels in more detail [4]; [5]; [10]; [14]. For instance, [10] concluded that vowels can be reliably distinguished if formant contour is considered, and such a model can help to separate those vowels whose formant values are very close and similar.

Many researchers have built on the direction model and reported that monophthongs can be characterised effectively when their formants are taken from three locations (at 20%, 50%, and 80%), e.g. [12], [16], among others. Another line of research building on the direction model takes the VISC measurement to an advanced level by measuring multiple samplings to represent detailed information of the entire formant trajectories (e.g., [1]; [6]; [11]; [14]; [22]). These studies suggest that using multiple measurements (as well as additional cues such as vowel duration, fundamental frequency [F0], and third formant frequency [F3]) are not only useful in

terms of describing monophthongs but also in terms of classifying monophthongal vowels using discriminant analysis.

Despite the importance of this research for more detailed crosslinguistic comparisons on vowel production, the role of dynamic correlates in L2 vowel production has not been fully investigated [24]. English and Arabic provide a good testing ground for the role of dynamic cues in vowel identity and L2 vowel learning, but we are not aware of research that has examined the dynamic properties of English monophthong vowels by Arabic L2 learners (see [2]; [3], among others, for examples of L2 production studies by Arabic L1 speakers using static models). This study constitutes the first examination of the dynamic production patterns of English vowels by Hijazi Arabic L2 (HA L2) learners and is also the first research contribution to examine L2 production comprehensively in terms of VISC. It examines the production patterns of Standard Southern British English (SSBE) vowels by HA L2 learners compared with English (NE) speakers and L1 patterns. It evaluates the relative importance of static and dynamic cues, particularly VISC, in describing and classifying SSBE vowel production by HA L2; it also explores to what extent vowel duration, F0, and F3 act as additional cues to classification accuracy.

2. METHODOLOGY

Data were collected from 60 participants (20 HA L1, 20 HA L2 learners, and 20 NE balanced by gender). To note, HA L2 learners were intermediate classroom English students in a foreign language setting and had no direct access to native L2 input. Recordings were made on a Roland Edirol R-09 recorder and Audio Technica Cardioid stereo microphone with a sampling rate of 44,100 Hz and 16-bit quantization. The target vowels were examined in a list with varied consonantal contexts. Each HA participants produced 48 words with one of eight HA vowels (/i iː eː a aː oː u uː/) and each L2 and NE participants produced 60 words with one of 10 English monophthong vowels (/ɪ iː ɛ ɔː ɑː ʊ uː æ ɒ ʌ/), with three repetitions (a total of 10,080 tokens). Vowel duration, the first three formant and F0 values were extracted from one location (50% for the static model) and multiple locations (two [20% and 80%]), three [20%, 50%, and 80%], and seven locations [20%, 30%, 40%, 50%, 60%, 70%, and 80%] for the dynamic models) over the course of the vowel duration.

Acoustic analysis was conducted using Praat [9]. The F0 settings were speaker-dependent and formant tracks were obtained using a 0.025 s window length, 50 Hz pre-emphasis, and a maximum frequency of 5,000 Hz for males and 5,500 Hz for females. For the offset model, we obtained the amount of a vowel's spectral changes by calculating the differences for all three formant and F0 values between the two vowel's positions. For the slope model, we obtained the vowel's rate of changes by using the offset value (see above) and then dividing them by the vowel duration. For the direction model, we computed the vowel's spectral shifts by tracking the first three formant and F0 values from two samples (for the two-point model), three samples (for the three-point model), and seven samples (for multiple points). All F0 and formant values were checked manually to ensure the accuracy of the results.

Two types of statistical techniques were used to evaluate the differences in the data—namely, pairwise comparisons (post-hoc tests) after a linear mixed effects regression, to determine the statistical significance of the study results [14]. Then a discriminant analysis (qda function) with a leave-one-out cross-validation, or 'jackknife' [5]; [16] was used as a classification tool to evaluate the extent to which the static and dynamic models and other acoustic feature sets (F0, F1, F2, F3, and vowel duration) improve vowel classification.

3. RESULTS

Beginning with the static model, HA L2 speakers were found to produce some SSBE vowels significantly (e.g., $p < 0.0001$) differently from the NE group, namely, /ʌ/, /ʊ/, /ɒ/, /æ/, /ɪ/, and /ɑː/ in terms of F0; /ʊ/ and /ɪ/ in terms of F1; /ɒ/, /ʌ/, /ɛ/, and /æ/ in terms of F2; and /ɛ/, /æ/, /ɔː/, /ʊ/, /uː/, and /ɒ/ in terms of F3. In contrast, HA and HA L2 groups produced vowels in a similar way (e.g., Figure 1 for F1/F2 estimates).

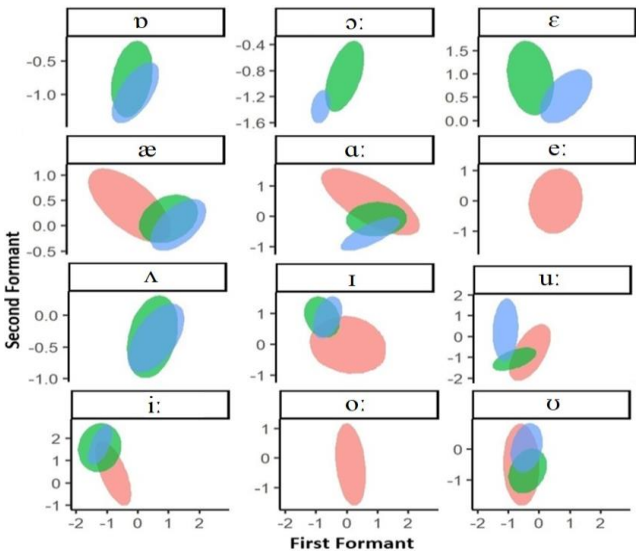

Figure 1: Scatter plot of the normalised midpoints (LOBANOV Z-scores) of the first two formant values of the vowels produced by HA (in red), HA L2 (in green), and NE (in blue) participants.

Regarding dynamic cues, particularly the offset model, HA L2 speakers were found to produce SSBE vowels with a great amount of spectral movement in terms of F1, F2, and F3, but fewer spectral changes in F0 (e.g., see Figure 2 for F2 changes). The results of the offset for HA and HA L2 speakers revealed that HA L2 speakers produced SSBE vowels with a similar amount of VISC as HA L1 speakers (e.g., a great amount of in F1, F2, and F3). HA L2 participants were found to produce most SSBE vowels significantly differently from the NE group: /ʊ/ in terms of F0; /ʌ/, /iː/, /ɔː/, /æ/, /ɛ/, and /ɑː/ in terms of F1; /ʊ/, /ɔː/, and /ɑː/ in terms of F2; and /ʌ/, /ʊ/, /æ/, and /ɑː/ in terms of F3.

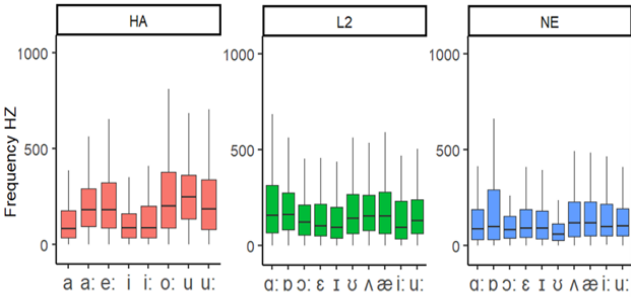

Figure 2: Box plot of the F2 offset model of the vowels produced by HA, HA L2, and NE participants.

In terms of the slope model, HA L2 participants were found to produce SSBE vowels significantly differently from the NE group, particularly for F0, and with some vowels in F1 (e.g., /uː/ and /ɛ/), F2 (e.g., /ɒ/ and /iː/), and F3 (e.g., /iː/). Once again, HA and HA L2 mostly had similar production of vowels, (e.g., see Figure 3 for F0 results).

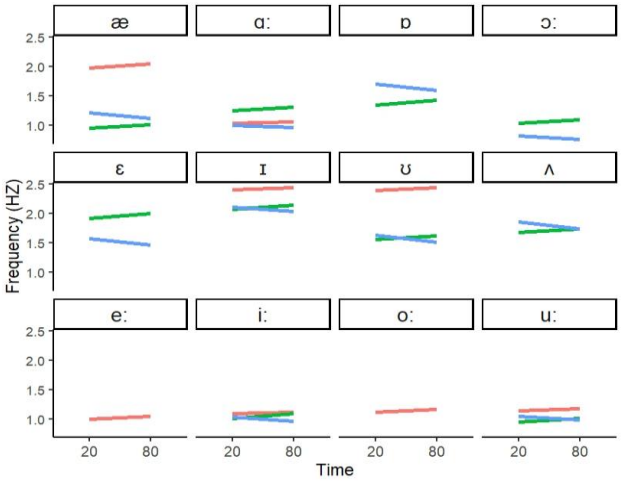

Figure 3: Results of the F0 slope model (measured at two points) of the vowels produced by HA (in red), HA L2 (in green), and NE (in blue) participants.

Regarding the direction of the two-point model, HA L2 was found to produce most SSBE vowels significantly dissimilar from the NE group (e.g., /ʌ/, /ʊ/, /ɒ/, /æ/, and /ɑː/ in terms of F0; /ɪ/ in terms of F1; /ɒ/ in terms of F2; and /ɔː/ in terms of F3) but similarly to their L1 production. The direction from three-point model showed the same results from the direction model from three-point measure, however we noted few more significant results between the production of HA L2 and NE (e.g., /ɪ/ in terms of F0, /ʌ/ and /ɛ/ in terms of F2, /uː/ and /æ/ in terms of F3. Using multiple points model (e.g., seven measurements) showed similar findings to the three-point model, and indeed, more significant results of the production of the L2. The most consistent result pertains to the HA L2 participants' production, which approximated the L1 (HA) participants in terms of the relative position of the vowels in the acoustic space, the direction, and the amount of the spectral change (e.g. Figure 4 for formant direction results).

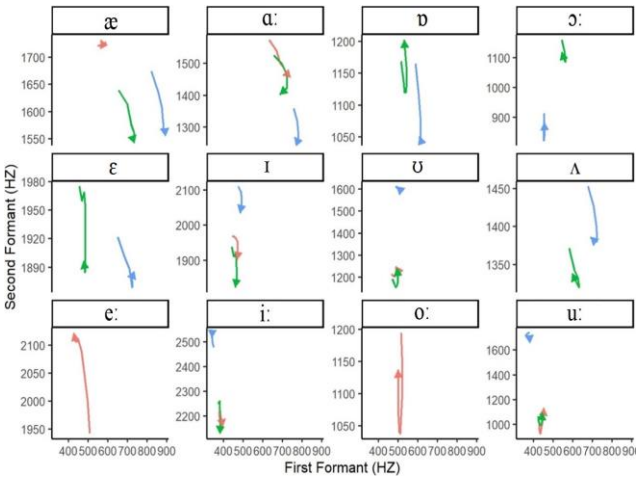

Figure 4: Vowel formant trajectories in F1-F2 space (measured at seven points) of the vowels produced by HA (in red), HA L2 (in green), and NE (in blue) participants; arrows represent the direction of formant movement.

In terms of classification. the QDA showed that using dynamic cues for all three groups (in particular, the seven-point model) with F0, F1, F2, and F3 (without and with the duration) resulted in the highest classification accuracy (with an average of 85% for HA; 80% for NE; and 61% for HA L2), followed by the three-point model (with an average of 76% for HA; 70.5% for NE; and 51% for HA L2), then by two-point model (with an average of 75% for HA; 68.5% for NE; and 50% for HA L2), and finally by the static model (with an average of 71% for HA; 65% for NE; and 48.5% for HA L2). With respect to the additional cues' results, vowel duration was found to play a significant role in the classification accuracy for HA and HA L2, while F0 was the most important additional cue for accurately classifying NE vowels. (see Table 1 for percentage improvement).

	The average improvement percentage		
	Vowel duration	F0	F3
HA L2	5.8	2.8	1.3
HA	11.8	2.1	1.6
NE	4.6	10	5.5

Table 1: The average improvement percentage in the discriminant classification accuracy of the three groups for all models with the addition of the vowel duration, F3, and F0 as additional cues.

4. DISCUSSION AND CONCLUSION

The data on the acoustic correlates of the production of SSBE by HA L2 speakers using the midpoint model showed that most of the SSBE vowels were produced differently from native-like productions, particularly for vowels that do not exist in the learner's L1 (HA), such as /ɔː/, /ʊ/, /ɒ/, /ɛ/, and /ʌ/, suggesting a difficulty in producing what are often referred to as 'new' vowels in the literature in a target-like manner [13]. Furthermore, the HA L2 static results for 'similar' vowels generally revealed that HA speaker produced many vowels of SSBE with similar patterns to their L1.

The effect of L1 on the performance of L2 was also found in the dynamic cues. For example, for the offset model, HA L2s were found to produce SSBE vowels with a degree of vowel-inherent spectral change that is similar to their HA L1 (e.g., greater spectral shifts in F1, F2, and F3). Such a result is expected due to speakers of low-density languages (e.g., HA) having more freedom and space to produce their vowels compared to high-density languages (e.g., SSBE; [7]; [17]; [19]). Similarly, the slope result showed that HA L2s had mostly similar slope values to their L1 (positive slopes in most cases related to faster spectral changes of HA monophthongal vowels during the vowel duration) and different from NE speakers. Interestingly each of the static, offset and slope models showed differences in terms of which vowels exhibited significant differences in their patterns between HA L2 and NE productions, and in which formants. This highlights the importance of considering measurement method before making firm conclusions about which vowels L2 speakers will find challenging to produce in a target-like manner. Conversely, all three methods stressed the influence of L1 on the production of L2.

Regarding the direction result (from two, three, or seven points), the HA L2 group once again produced most SSBE vowels in similar patterns of formant trajectories to those of their L1 and different from NE speakers. While this supported results from the other models, the data of the direction model revealed that using dynamic measurements provided significantly more information and differences about the production of vowels than the static model did. Moreover, more measuring points from the vowel duration provided richer information about the fuller extent of the vowel spectral changes that might remain unnoticed when formant values are taken from fewer locations. This result supports the necessity of investigating monophthongal vowels by using dynamic measurements (multiple points) [1]; [14]; [22], among others. Overall, such findings confirm that the VISC models could be used as another perspective to examine more closely the extent to which L2 learners are influenced by their L1 [20].

With respect to the classification accuracy of vowels, the QDA results suggest that extracting two measurement points (or more) from the vowel formants shows notable improvements compared with taking a single point. This finding provides support for the dynamic approach [4]; [5]; [12]; [16]. Among the dynamic measurements, the seven-point measure is the most accurate for classifying the vowels' production for HA L2. Such a finding indicates that entire trajectory of the formants might contain additional vowel identity information over and above that captured in two- or three-point model [20]. We also noted that the QDA outcome's average rate of the classification accuracy for HA L2s was very low compared to that of NE speakers which is expected due to the noticeable variations in producing SSBE vowels in all acoustic models used. Despite the efficiency of the F1 and F2 values in identifying vowels, this study highlights the role of additional cues in providing more insights and increased separation of vowels. For example, HA L2s relied mostly on vowel duration (similarly to their L1), and this can be explained by considering the phonological role of vowel duration as a cue to distinguishing short and long vowels in Arabic vowels [4]; [5]; [8]; [18]. In addition, the use of F0 and F3 by each of HA L2 participants in SSBE vowel production, exhibited influence from the speakers' L1. Taken Together, these findings reveal that cues to vowel identification are not expressible in one time location and that the transitional changes perform significant functions in terms of describing and classifying monophthongal vowels [4]; [5]; [6]; [15]; [16]; [20]; [21]. These can be missed when crosslinguistic vowel comparisons are reduced to single points, and in turn concealing potentially important differences between L1 and L2 speakers' production of these vowels.

To sum up, our findings are consistent with dynamic theories of vowels and heed recent calls to expand the sociophonetic toolkit beyond single-point measures and beyond the first two formants to represent more information and reflect the dynamic complexity of actual vowel movement. As quoted from [25], 'traditional two-dimensional vowel charts are not sufficient, and … a deeper understanding is gained when dynamic patterns of formant movement are documented'; we believe we underscored this point in this paper.

5. ACKNOWLEDGMENTS

This work was supported by Taibah University to the first author (WA) and partially supported by the French Investissements d'Avenir—Labex EFL program (ANR-10-LABX-0083), contributing to the IdEx Université Paris Cité—ANR-18-IDEX-0001 to the second author (JAT). We thank all of our subjects who participated in this study.

6. REFERENCES

[1] Adank, P., Van Hout, R., and Smits, R. 2004. An acoustic description of the vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical Society of America* (*J. Acoustic. Soc. Am*), 116(3), 1729-1738.

[2] Ali, E. 2013. Pronunciation problems: Acoustic analysis of the English vowels produced by Sudanese learners of English. *International Journal of English and Literature*, 4(10), 495-507.

[3] Almbark, R. 2014. Perception and production of SSBE vowels by foreign language learners: Towards a foreign language model. In *Proceedings of the International Symposium on the Acquisition of Second Language Speech: Concordia Working Papers in Applied Linguistics*, 5, Montreal, Canada, 3-18.

[4] Almurashi, W., Al-Tamimi, J., and Khattab, G. 2019. Static and dynamic cues in vowel production in Hijazi Arabic. In *Proceedings of the 19th ICPhS*, Melbourne, Australia, 3468-3472.

[5] Almurashi, W., Al-Tamimi, J., and Khattab, G. 2020. Static and dynamic cues in vowel production in Hijazi Arabic. *J. Acoustic. Soc. Am*, 147(4), 2917-2927.

[6] Al-Tamimi, J. 2007. Static and dynamic cues in vowel production: a cross dialectal study in Jordanian and Moroccan Arabic. In *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, Saarbrücken, Germany, 541-544.

[7] Al-Tamimi, J., and Ferragne, E. 2005. Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. In *Proceedings of the 9th European Conference on Speech Communication and Technology*, Lisbon, Portugal, 2465-2468.

[8] Al-Tamimi, J., and Khattab, G. 2015. Acoustic cue weighting in the singleton vs geminate contrast in Lebanese Arabic: The case of fricative consonants. *J. Acoustic. Soc. Am*, 138(1), 344-60.

[9] Boersma, P., and Weenink, D. 1992–2022. Praat: Doing phonetics by computer.

[10] Chladkova, K., and Hamann, S. 2011. High vowels in Southern British English: /u/-fronting does not result in merger. In *Proceedings of the 17th ICPhS*, Hong Kong, China, 476-479.

[11] Darcy, I., and Mora, J. C. 2015. Tongue movement in a second language: the case of Spanish/ei/-/e/ for English learners of Spanish. In *Proceedings of the 18th ICPhS*, Glasgow, UK.

[12] Ferguson, S. H., and Kewley-Port, D. 2002. Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *J. Acoustic. Soc. Am*, 112(1), 259-271.

[13] Flege, J., and Bohn, O. S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (3-83). Cambridge University Press.

[14] Fox, R. A., and Jacewicz, E. 2009. Cross-dialectal variation in formant dynamics of American English vowels. *J. Acoustic. Soc. Am*, 126(5), 2603-2618.

[15] Hillenbrand, J., Clark, M., and Nearey, T. M. 2001. Effects of consonant environment on vowel formant patterns. *J. Acoustic. Soc. Am*, 109(2), 748-763.

[16] Hillenbrand, J., Getty, L. A., Clark, M., and Wheeler, K. 1995. Acoustic characteristics of American English vowels. *J. Acoustic. Soc. Am*, 97(5), 3099-3111.

[17] Jin, S. H., and Liu, C. 2013. The vowel inherent spectral change of English vowels spoken by native and non-native speakers. *J. Acoustic. Soc. Am*, 133(5), 363-369.

[18] Khattab, G. 2007. A phonetic study of gemination in Lebanese Arabic. In *Proceedings of the 16th ICPhS*, Saarbrücken, Germany, 153-158.

[19] Meunier, C., Frenck-Mestre, C., Lelekov-Boissard, T., and Le Besnerais, M. 2003. Production and perception of vowels: Does the density of the system play a role? In *Proceedings of the 15th ICPhS*, Barcelona, Spain, 723-726.

[20] Morrison, G. S., and Assmann, P. 2013. *Vowel inherent spectral change*. Springer Science and Business Media.

[21] Nearey, T. M., and Assmann, P. F. 1986. Modeling the role of inherent spectral change in vowel identification. *J. Acoustic. Soc. Am*, 80(5), 1297-1308.

[22] Neel, A. T. 2004. Formant detail needed for vowel identification. *Acoustics Research Letters Online*, 5(4), 125-131.

[23] Peterson, G. E., and Barney, H. L. 1952. Control methods used in a study of the vowels. *J. Acoustic. Soc. Am*, 24(2), 175-184.

[24] Rogers, C.L., Glasbrenner, M.M., DeMasi, T.M., and Bianchi, M. 2013. Vowel inherent spectral change and the second-language learner. In G. S. Morrison, and P. F. Assmann (Ed.), *Vowel inherent spectral change* (231-259). Springer, Berlin.

[25] Schwartz, G., Aperliński, G., Kaźmierski, K., and Weckwerth, J. 2016. Dynamic targets in the acquisition of L2 English vowels. *Research in Language*, 14(2), 181-202.