Psicológica (2007), 28, 167-176.

Stress placement and word segmentation by Spanish speakers

Juan Manuel Toro-Soto[1♦], Antoni Rodríguez-Fornells[2,4]

& Núria Sebastián-Gallés[3, 4]

[1] *International School for Advanced Studies (SISSA/ISAS), Trieste*
[2] *Institució Catalana de Recerca i Estudis Avançats (ICREA)*
[3] *GRNC, Parc Científic de Barcelona, Universitat de Barcelona*
[4] *Departament de Psicología Básica, Universitat de Barcelona*

Several studies have shown that the stress pattern of one's native language is applied to new linguistic stimuli. Regarding the segmentation of artificial synthesized speech, this idea has been supported by experiments with languages where the stress pattern coincides with word boundaries (i.e. English, Finnish and Dutch). In this study, we present data on speech segmentation with native Spanish speakers whose stress pattern would mark the penultimate syllable of words. Results show that to stress the middle syllable of trisyllabic words in an artificial speech stream does not facilitate segmentation as would be predicted. Possible explanations of these results are explored as related to the interaction of statistical and stress cues in speech segmentation.

Research on speech segmentation has indicated that, when facing the problem of segmenting a novel speech stream, be it an unfamiliar or artificial language, the individuals use the patterns from their own maternal language and apply them to the new stimuli (Cutler et al, 1986; Johnson & Jusczyk, 2001; Sanders & Neville, 2000; Vroomen et al. 1998).

Mainly, these studies try to understand how different cues present in the speech signal such as language rhythm, phonotactics, vowel harmony, or stress patterns are used by a learner to segment the words composing the language. Among these cues, the statistical regularities in speech have been widely studied (i.e. Saffran, 2002; Saffran, Aslin & Newport, 1996; Saffran,

♦ Acknowledgments: This study was supported by the James S. McDonnell Foundation (JMSF 20002079) and the Spanish Ministerio de Educación y Ciencia (with EC Fondos FEDER: SEJ2004-07680-C02-01/PSIC and AP2000-4164). We thank Elena Azañon for her help in running the experiment, and members from the GRNC for valuable comments on the study. Correspondence should be addressed to: Juan Manuel Toro Soto. Cognitive Neuroscience Sector. SISSA/ISAS. Via Beirut, 2/4. 34014 Trieste, Italy. E-mail: jmtoro@sissa.it

Newport & Aslin, 1996; Newport & Aslin, 2004; Peña, Bonatti, Nespor & Mehler, 2002). These studies have demonstrated that human adults and infants can effectively extract on-line regularities over a speech stream by the computation of transitional probabilities among the items composing the words in the stream. Also, they have established a reliable methodology for studying how different acoustic cues interact when a listener faces the problem of speech segmentation (Vroomen, Toumainen & de Gelder, 1998) using mainly artificial synthesized stimuli.

Regarding the interactions between different cues during the segmentation of speech by adults, Saffran, Newport & Aslin (1996) found that the lengthening of the final syllable improved performance in a segmentation task. This facilitatory cue was thought to be available to listeners across many linguistic environments, since, in natural utterances, the final syllable tends to be lengthened with respect to the others. Also, in Experiment 3 of their study, Vroomen et al. (1998) showed that native speakers of Finnish and Dutch performed better in a segmentation task when initial stress was added to the words composing an artificial speech stream. Nevertheless, native speakers of French did not benefit from this modification. The authors' explanation of these results pointed to differences in the stress pattern of Finnish, Dutch, and French. Finnish and French are languages with fixed lexical stress. Multi-syllabic content words in Finnish always have stress on the initial syllable, while in French stress is on the final syllable. Additionally, in Dutch, most words have stress on the initial syllable, although there are a few variations (Vroomen et al. 1998). So, the authors concluded that a listener's performance in a segmentation task improved when the phonological cues in the stream matched those of their maternal language (pp. 145). However, the results of French participants could also be explained by their difficulties in perceiving lexical stress, as shown in the studies of Dupoux and collaborators (Dupoux, Christophe, Sebastián-Gallés & Mehler, 1997; Dupoux, Peperkamp & Sebastián-Gallés, 2001). It is not surprising then that French speakers did not take advantage of the stressed cue.

Results obtained with infants also suggest an early sensitivity to the predominant stress pattern in speech segmentation. Seven-and-a-half month-old English learners can segment strong-weak, but not weak-strong, bisyllabic words out of a stream, which coincides with the most common pattern in English. It is not until 10.5 months of age that infants can integrate other stress patterns, such as weak-strong (Jusczyk, Houston & Newsome, 1999). Regarding the use of conflicting cues in speech segmentation, Johnson & Jusczyk (2001), using speech streams composed by the same trysillabic words as in the Saffran, Aslin & Newport (1996) experiments, observed that 8-month-old infants raised in an English-

speaking environment relied more heavily on stress cues than on statistical cues when the stress was put on the initial syllable of the word. So the notion that language specific stress patterns were applied to new linguistic stimuli was supported. Finally, Thiessen & Saffran, (2003) demonstrated that by 9 months of age, English-learning infants still rely more on a trochaic stress pattern than on statistical cues when segmenting an artificial language, which replicates the results by Johnson & Jusczyk (2001). They also showed that this preference seems to develop with age, since by 7 months of age infants rely more heavily on statistical cues.

To our knowledge, the research on the interaction of different cues when confronting a segmentation task has been mainly done over languages with stress primarily on the initial syllable of a word, such as English, Finnish and Dutch; or languages that exclusively stress the final syllable, such as French. But it is yet unknown what would happen in the case of languages without a stress initial pattern, in particular in languages with stress medial patterns. That is, languages in which the stress does not coincide with one of the boundaries of the word, and whose native speakers actively use stress information in lexical search, as it is the case of Spanish (e.g. Soto-Faraco, Sebastián-Gallés & Cutler, 2001).

In Spanish all multisyllabic words have one syllable marked for primary stress. Penultimate stress is predominant (in 75% to 80% of the words, the syllable marked for primary stress is the second-to-last syllable; Harris, 1983; Quilis, 1984). This is also the case for trisyllabic words, for which Spanish has a stress-medial pattern (Navarro, 1966). The percentage of trisyllabic words in Spanish that stress the medial syllable is 73.52% (as calculated over the LEXESP database that contains over 5,020,930 word items; Sebastián-Gallés, Martí, Carreiras & Cueto, 2000). For a Spanish speaker then, it is very likely that the preferred stressed syllable in a trisyllabic stimulus would mark the middle section of the word, and not one of its boundaries.

If the hypothesis that stress patterns from a participant's native language are applied to novel stimuli is correct, then stressing the medial syllable in a speech stream should facilitate the segmentation process in Spanish speakers, since Spanish has a stress medial pattern in trisyllabic words. To test this hypothesis, an experiment using artificial languages with different stress patterns was run. We created three artificial streams with identical structure to the ones used by the Saffran, Aslin & Newport (1996) to study speech segmentation with statistical information. In these three streams, either the first, second or third syllables of every word would be stressed, thus yielding different stress patterns. As a baseline, a fourth stream with flat stress was also created. Using this stream, we wanted to be

sure that participants could reliable segment out the words using only statistical cues.

EXPERIMENT

METHOD

Participants. The participants were 85 undergraduate Psychology students at the Universitat de Barcelona (n=17 in each condition). They were native speakers of Spanish. No hearing deficits or problems in language acquisition were reported by any participant. They were given course credits for their participation on the study.

Stimuli and apparatus. Four artificial languages that differed only in the stress pattern on their words were used. Languages were formed by four trisyllabic nonsense words (hereafter referred to as "words"; *tupiro, golabu, bidaku, padoti*). These four words were concatenated in a stream following the same order as in the above mentioned study, thus there were not any immediate repetitions of words. Statistical regularities were controlled, so that the probability for a given syllable to be followed by another given one within a word 1.0, while it was 0.33 between words. The sequence of words was synthesized using the text-to-speech MBROLA software (Dutoit et al. 1996) with a Spanish male diphone database[1] at 16 kHz. The most important feature of the speech stream was that there were no acoustic markers between words, so the only cues available to segment out the words were the statistical ones (flat stress language).

To create the stressed languages, one syllable of each word was modified. This modification was done by augmenting a given syllable's pitch by 20 Hz. To create the stress-initial, the stress-medial language, and the stress-final language, the first syllable, the second syllable, or the third syllable of each word was changed. To control for any undesirable effect of rhythm by introducing stress changes every three syllables, a fifth language was constructed using the same syllables as in the previous one. But this time they were arranged in a completely random order, so there would not be any reliable statistic cue to group syllables, just that every three syllables, one was stressed. This language was synthesized using the same procedure, and lasted the same as all the other languages. So, in total, we created three stressed languages (stress-initial, -medial, and –final), one language with flat stress, and one language with a random organization of syllables.

[1] available at http://tcts.fpms.ac.be/synthesis/mbrola.html

To assess the segmentation of the stream, eight test items were created using the same synthesis procedure used in the languages. There were four words and four part-words. None of them was stressed. Words were the same four items that composed the stream. Part-words were made by the concatenation of one syllable of a word and two syllables of another one, and were *tibida, kupado, rogola, butupi.*

Procedure. The experiment was run individually in a sound-attenuated booth, on a PC computer using the EXPE programming language (Pallier, Dupoux, and Jeannin, 1997). Stimuli were played throughout Sennheiser (HMD224) headphones connected to the computer, via a Proaudio Spectrum 16 soundcard. Each participant was exposed to one of the languages. After 7 minutes of listening to the language, they were given a 2 alternative forced choice (2AFC) test. They were presented with word - part-word pairs, with an interval of 500 ms between each item, and had to answer by pressing either a "1" or "2" key in a response box indicating whether the first or the second trisyllabic groupings was more likely to be a word in the language. The next test pair was presented either when the participant pressed a response key, or when 5 seconds had elapsed after the presentation of the previous test pair. There were eight test trials, each one consisting on the presentation of a test pair in which the order of word – part-word was balanced.

RESULTS

Results for the five conditions are shown in Figure 1. An ANOVA with condition as the between-subjects variable revealed significant differences in percentage of correct responses ($F(4, 80)=10.851$, $p<0.001$). Post-hoc analysis showed two homogeneous subsets with no differences within each subset. The first one was composed of the *stress-flat, stress-initial,* and *stress-final* conditions. Percentage of correct responses for subjects presented with the stress-flat language was 70.59% ($SD= 18.19$). This percentage was significantly different from chance ($t(16)= 4.667$, $p<0.001$), demonstrating that subjects could segment the words from the speech stream using statistical cues alone. This result resembles previously reported ones for adult subjects with part-word foils (65% in Saffran, Newport & Aslin, 1996), suggesting a strong effect for segmentation with statistical cues alone. For subjects presented with the stress-initial language the percentage of correct responses was 71.32% ($SD= 16.39$), which also differed from chance ($t(16)= 5.362$, $p<0.001$). Finally, for subjects presented with the stress-final language, percentage of correct responses was 76.47% ($SD=16.46$), and differed from chance too ($t(16)= 6.628$, $p<0.001$). Comparisons in responses for stress-flat and stress-initial

languages with the stress-final language yielded non-significant results (for all comparisons, $t<1$). So, stressing the first or the last syllable in the words composing the stream did not have any effect in the segmentation results by Spanish speakers, when compared with the results from the flat-stress language.

The second subset was composed by the stress medial plus the random condition. Percentage of correct responses for subjects presented with the random language was 49.26% (SD=18.99). This result did not differ from chance ($t(16)$= -0.160, p=0.875). So, as expected, when the stream contained no reliable statistical cue to group syllables, participants were at chance during the test. This result implies first, that there are no a priori preferences for either type of test items (words or part-words) in the present experiments, and second, that introducing a stress change every three syllables has no grouping effect by itself during the segmentation of the stream. Concerning the results of the *stress-medial* condition, subjects had a percentage of correct responses of 41.17% (SD=25.68). This result was also not significantly different from chance ($t(16)$= -1.417, p=0.176). Not only the accentuation of the middle syllable in the words did not help the subjects to obtain higher scores, but with this stress pattern the subjects' performance dropped to chance levels.

Taken together, these results point in the opposite direction as that originally predicted. When stress fell in the middle syllable of trisyllabic words it did not facilitate the segmentation of the speech stream by Spanish listeners. Even more striking is the fact that performance dropped to chance levels for the stress-medial language, suggesting a possible conflict between statistical and stress cues. While statistical cues would mark the boundary of a word at the first and third syllables, stress would make more salient the second one, possibly drawing attention and processing resources towards it. Results suggest that this conflict was not present in stress-initial and stress-final languages, since performance with them was not significantly different from performance with the stress-flat language.

GENERAL DISCUSSION

Contrary to what had been hypothesized, stressing the middle syllable of words composing an artificial speech stream did not help Spanish speaking participants on a segmentation task. Even more, it made performance drop to chance levels. Also, stressing the initial syllable did not improve performance either, as it did in studies whose participants native language was mostly stressed on the initial syllable (e.g. Vroomen et al. 1998). For Spanish speakers, the results for the stream without stress,

and the stream with stress on either the initial or the final syllable, were the same.

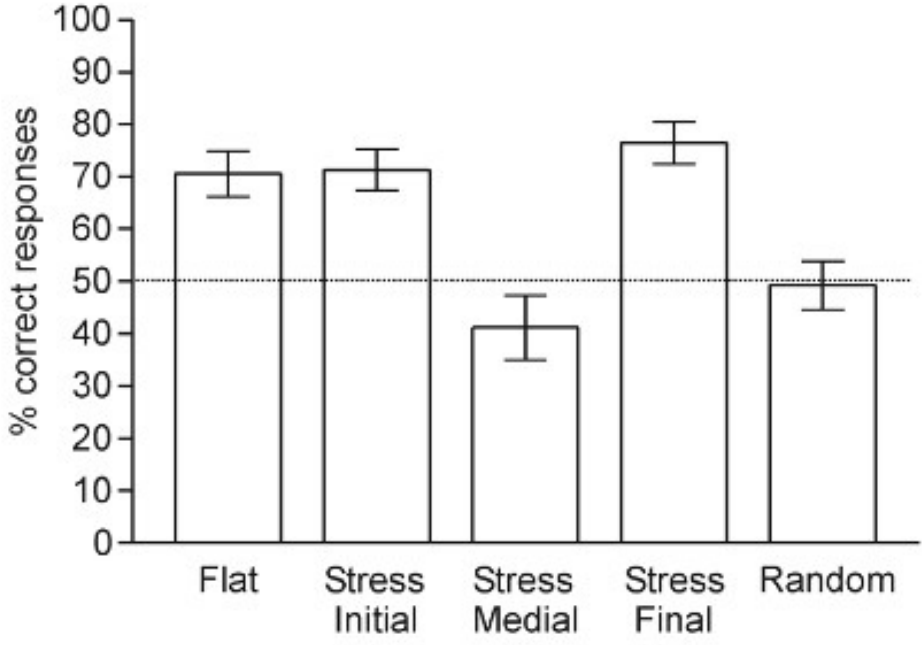

Figure 1. Mean percentage (and standard errors) of correct responses for the different experimental conditions in the present experiment.

Does this mean that listeners do not apply the stress pattern of their native language to new speech stimuli? When stress marks one of the boundaries of words, as in the case of stress initial languages (English, Finnish) or stress final languages (French), it helps the listener to segment out the words. But when stress falls on the middle syllable of trisyllabic words, it seems to conflict with statistical information that marks the beginning and ending of words. Maybe, stressing the initial syllable makes the listener to pay more attention to the beginning of the word. In this way it would not be a conflicting cue with respect to statistical information for Spanish speakers, and therefore their performance would be equivalent to that with flat stress speech streams, as was observed in this study. As for the effect of stressing the middle syllable, it is likely that it brings the attention of the listener to a syllable that does not mark a limit of the word, and this is reflected in the test with chance levels of performance.

Unfortunately, there are not extensive studies that describe the mechanisms involved in the segmentation of continuous speech by Spanish

speakers. There are, however, indications that stress cues are indeed taken into account during segmentation in Spanish (Sebastián-Gallés & Costa, 1997), and that they play an important role in the activation of lexical entries (Soto-Faraco, Sebastián-Gallés, & Cutler, 2001). Recent models of speech segmentation propose that there are different cues contributing to speech segmentation (from lexical to suprasegmental; Mattys, White, and Melhorn, 2005), and that the weight, or relative importance, of them may change depending of specific properties of different languages. So, for example, stress and syllabic information may contribute differently to the segmentation of speech in languages such as Spanish and English. While stress would have more importance in English, the syllable would have it in Spanish. It is therefore an open question the extent to which the present results would generalize to other languages, as importance of stress changes may vary across them.

As mentioned earlier, in Spanish around 75% of the words the second-to-last is the primary stressed syllable. One could suggest that this percentage is not enough for listeners to actually apply this pattern to new artificial stimuli. Nevertheless, in English, around 63% of the words have primary stress in the first syllable (taking also into account monosyllabic words; Cutler, 1990), and it has been shown that speakers of English apply their native language stress pattern in segmentation experiments (e.g. Sanders & Neville, 2000; Vroomen et al. 1998). This percentage of words in Spanish should be sufficient to create a stress pattern that native speakers would apply to new stimuli, such as the one used in the present study. Anyway, variability in the position of stress in Spanish, would not predict that stressing the middle syllable of words in an artificial speech stream would bring segmentation performance to chance levels.

It is thus possible that salience of certain statistically-coherent groupings is increased by subtle phonetic modifications, as the ones used in this study. To stress the middle syllable may bring the listener to extract regularities that may not necessarily coincide with the ones for segmenting the trisyllabic words out of the stream. Thus, the effect of stress found in previous studies using English speakers would likely reflect an "edge effect". In languages that stress the first syllable of words, such as English, the perceptual properties of the stimuli would coincide with the linguistic ones, so both perceptual salience and lexical stress would mark the beginning of words. In this way, both language-specific and general perceptual information would coincide, allowing for easier word segmentation. In any case, this is an interesting possibility that further research with other languages should investigate. As for the results presented here, they suggest that, under certain conditions, perceptual

salience may override language-specific stress patterns during a word segmentation task.

RESUMEN

Posición del acento y segmentación de palabras en hablantes de Español. Varios estudios han mostrado que el patrón acentual de la lengua materna se aplica a estímulos lingüísticos novedosos. En el campo de la segmentación del habla sintetizada, esta idea ha recibido apoyo de experimentos con lenguajes en los cuales el patrón acentual coincide con los límites de las palabras (p.e. Inglés, Finlandés y Holandés). En este estudio, presentamos datos sobre la segmentación del habla en hablantes de Español, cuyo patrón acentual tiende a marcar la penúltima sílaba de las palabras. Los resultados muestran que acentuar la sílaba del medio en las palabras trisilábicas de un flujo de habla artificial no facilita su segmentación, tal y como se podría predecir. Se exploran pues las posibles explicaciones de estos resultados, en la medida en que se relacionan con la interacción de claves estadísticas y acentúales durante la segmentación del habla.

REFERENCES

Aslin, R., Saffran, J., & Newport, E. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9,* 321-324.

Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In Altmann, G. *Cognitive models of speech processing: Psycholinguistic and computational perspectives.* (pp. 105-121). The MIT Press. Cambridge.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25,* 385-400.

Dupoux, E., Christophe, P., Sebastián-Gallés, N., & Mehler, J. (1997). A distressing deafness in French. *Journal of Memory and Language, 36,* 406-421.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness". *Journal of the Acoustical Society of America, 110,* 1606-1618.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O. (1996). *The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes*. ICSLP, Philadelphia.

Harris, A. (1983). *Syllable structure and stress in Spanish: A nonlinear analysis*. The MIT Press. Cambridge.

Johnson, E., & Jusczyk, P. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44,* 548-567.

Jusczyk, P., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39,* 159-207.

Mattys, S., White, L., & Melhorn, J. (2005). Integration of multiple segmentation cues: a hierarchical framework. *Journal of Experimental Psychology: General, 134,* 477-500.

Navarro, T. (1966). *Estudios de fonología española*. Las Américas. New York. [*Studies of Spanish Phonology*].

Newport, E., & Aslin, R. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48,* 127-162.

Pallier, C., Dupoux, E., & Jeannin, X. (1997). EXPE: An expandable programming language for on-line psychological experiments. *Behavior Research Methods, Instruments, and Computers, 29*, 322-327.

Peña, M., Bonatti, L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298,* 604-607.

Quilis, A. (1984). *Métrica española.* Ariel. Barcelona. [*Spanish Metric*].

Saffran, J. (2002). Constraints on statistical learning. *Journal of Memory and Language, 47,* 172-196.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-olds infants. *Science, 274,* 1926-1928.

Saffran, J., Newport, E., & Aslin, R. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35,* 606-621.

Sanders, L. & Neville, H. (2000). Lexical, syntactic, and stress-pattern cues for speech segmentation. *Journal of Speech, Language, and Hearing Research, 43,* 1301-1321.

Sebastián-Gallés, N., & Costa, A. (1997). Metrical information in speech segmentation in Spanish. *Language and Cognitive Processes, 12,* 883-887.

Sebastián-Gallés, N., Martí, M., Carreiras, M., & Cuetos, F. (2000). *LEXESP: Léxico informatizado del español.* Ediciones Universitat de Barcelona. Barcelona. [*LEXESP: Spanish informatized lexic*].

Soto-Faraco, S., Sebastián-Gallés, N. & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45*, 412-432.

Thiessen, E. & Saffran, J. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology, 39,* 706-716.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language, 38,* 133-149.

(Manuscript received: 2 October 2006; accepted: 5 December 2006)