The interplay of stress saliency and word beginning saliency: an experimental study

Article

Published Version

Cilibrasi, L. and Stojanovik, V. (2020) The interplay of stress saliency and word beginning saliency: an experimental study. Linguistica Pragensia, 30 (2). pp. 113-126. ISSN 0862-8432 doi: https://doi.org/10.14712/18059635.2020.2.1 Available at http://centaur.reading.ac.uk/95541/

Identification Number/DOI: https://doi.org/10.14712/18059635.2020.2.1 <https://doi.org/10.14712/18059635.2020.2.1>

Publisher: De Gruyter

The interplay of stress saliency and word beginning saliency: an experimental study

Luca Cilibrasi (Charles University, Prague) — Vesna Stojanovik (University of Reading)

ABSTRACT

There is a robust amount of evidence (mostly from English) suggesting that, while listening to speech, the initial part of words is scrutinised with more attention. Similarly, data suggests that stressed syllables are processed with more precision than unstressed syllables. How do these two kinds of saliencies interact? In this experimental study, the issue was investigated in a group of Italian speakers. Participants were presented with minimal pairs of nonwords differing in one individual phoneme (and specifically one trait, voicing). Nonwords were created as to contain phonological clusters in either an initial or medial position, and, similarly, stress was placed in either initial or medial position. Results show that when the clusters were in word medial position, there was a large effect of stress, with stressed syllables being recognised with greater accuracy. When the clusters were in initial position, instead, accuracy was at an intermediate level and we did not observe any effect of stress. The result is discussed in relation to previous literature addressing these phenomena in English.

KEYWORDS

English phonology, Italian phonology, nonwords, perception, saliency, stress, word beginnings

DOI

https://doi.org/10.14712/18059635.2020.2.1

1 INTRODUCTION

When, in our everyday life, we speak, the acoustic conditions are not always ideal, and we often find ourselves trying to parse speech in noisy environments. Sometimes the speech we are trying to interpret overlaps with that of other people, sometimes noises of cars or animals or music can cover the voice of whom we are listening to. Classic studies addressing the impact of noise on word recognition show that placing short noises on words has a different effect depending on which position of the word is noised (Walley 1988). While noises in word final positions do not lead to substantial difficulties in word recognition, noises in word initial positions are detrimental for our access to the lexicon (Nooteboom 1981). Findings of this kind have a very intuitive (and plausible) interpretation: word initial positions are somehow more important in speech processing and lexical access.

There is a considerable amount of linguistic and psycholinguistic evidence suggesting that different positions in the word are processed differently. Word initial syllables are described as strong, since they license a large number of contrasts and resist reduction (Beckman 2013, Smith 2004), and non-initial syllables are described as weak, since they license a smaller number of contrasts and tend to reduction (Beckman 2013, Harris 2011). Word beginnings play a major role in lexical access (Zwitserlood 1989, Marslen-Wilson & Zwitserlood 1989, Pitt & Samuel 1995) and are more likely to be remembered in the *tip of the tongue* phenomenon (Browman 1978). This pattern is sometimes described as a word beginning saliency principle (Beckman 2013). The principle operates in typical and clinical populations. For instance, in children with language impairment, clusters in initial syllables are repeated more accurately than clusters in word medial positions (Marshall and var del Lely 2009).

The word beginning saliency principle proposed by Beckman (2013) is described as a general phonological pattern that applies to the perception of any spoken material in both typical and atypical populations across languages. This framing predicts that there will be better accuracy in the detection of complex phonological clusters in word beginnings compared to the detection of clusters in word medial syllables in speakers of any language. This study tests this prediction on Italian speakers.

There are good reasons to test this prediction on Italian. Most of the data presented above comes from studies on English. One possible explanation for such results (alternative to the saliency principle) is linked to the nature of English stress. In English, stress can occur in all different positions in words, such as initial, medial or final syllables (Roach 2000). However, the distribution of stress in English is not random and it reveals relatively consistent patterns.

The classic view posits a general "English stress rule", claiming that penultimate stress is a default option (Hayes 1982) and words that do not respect this pattern are seen as exceptions. Recent studies capture variation in a more analogical fashion. First, these studies emphasize that English stress is related to syllable weight. Weight is a phonological property that reflects the number of segments in a given syllable. Syllables with long vowels and/or complex codas are heavier than syllables with short vowels and no coda (though the exact calculation of weight is disputed). Stress tends to be attracted to heavy syllables, and, in English polysyllabic words, the most common outcome is a heavy penultimate syllable (Domahs, Plag and Carroll 2014).

Second, it has been calculated that in English the majority of words receive stress on the first syllable. This is consistent with the stress assignment preferences just discussed, since it is a consequence of the fact that in English most words are monosyllabic or disyllabic, and thus stress either occurs on the only syllable available, or it occurs on the first one, which is also the penultimate (Jusczyk, Cutler & Redanz 1993). As Domhas, Plag and Carroll (2014: 60) put it:

> "It has been uncontroversially proposed that German, Dutch, and English are trochaic languages that are regularly stressed on the penult, if the final syllable is reduced. Native words in West Germanic languages are frequently bisyllabic with a reduced final syllable, automatically leading to penultimate stress."

Considering the frequency of the stress in initial position, and considering the importance of stress as a saliency factor (Marshall & van der Lely 2009), it may be argued that in English initial positions are salient also from the stress point of view (and not only because they are word beginnings). This is, however, not the case for all languages. In Italian, for example, most of the words receive stress on the penultimate syllable, and the number of words with more than three syllables is considerably larger than in English (Guasti 2004, Hayes 2012). Though underlining that this is an approximation, Krämer offers a numerical estimate (2009: 161):

> "The general estimate is that penultimate stress occurs in around 70–80 per cent of all lexical items in Italian, antepenultimate stress in around 20 per cent, and final stress in around 2 per cent."

The fact that languages have different prototypical stress patterns is considered a fundamental cue for lexical segmentation during early infancy. As soon as children realize the nature of the stress pattern in their language, they can determine where words start and finish in the phonetic wave (Guasti 2004, Mehler, Dupoux, Nazzi & Dehaene-Lambertz 1996, Saffran, Aslin & Newport 1996, Echols 1996, Ambridge & Lieven 2011, Dehaene 2009, Guttorm et al. 2010, Benasich & Tallal 2002, Männel & Friederici 2008). Stress patterns, thus, are very important in speech perception, and may act against the word beginning saliency principle (Beckman 1998) in a language such as Italian. Thus, given the nature of its pattern, Italian appears as an ideal language to investigate the relation between the word beginning saliency principle (Beckman 1998) and stress. In Italian, the stress pattern saliency may operate in competition with the word beginning saliency principle, offering a peculiar context for the investigation of the relation between the two saliencies. In English, on the contrary, the stress pattern saliency may operate in concert with the word beginning saliency principle, offering a blurrier picture. Experimental data on Italian may be helpful in explaining how the two kinds of saliencies interact.

2 METHODS

This section contains a thorough description of the methods, divided in the following sub-sections: ethical concerns, calculation of the sample size, description of the participants, description of the stimuli, procedure, and scoring.

2.1 ETHICS, RECRUITMENT AND CONSENT

The current study was approved by the University of Reading Research Ethics Committee and it was given favourable opinion. Adult native speakers of Italian were recruited by contacting the University of Reading Italian Society. The Society is a non-profit association of Reading University students who are interested in Italian language and culture. Members can be of any nationality. The society put the researcher in contact with a group of Italian students who were spending the academic year (2013/2014) in Reading as part of the Erasmus Programme. The students were informally approached during society meetings, where the researcher distributed information sheets about the present study with contact details. If interested in taking part, the students contacted the researcher and a date and a time for testing were agreed. Testing was conducted in a suitable laboratory at the School of Psychology and Clinical Language Sciences. Before testing, students were given time to reread the information sheet and if happy to proceed with the study, they were asked to sign a consent form. Participation was voluntary and students did not receive any reward for their participation in this study. After the testing, the students, if interested, were provided with a precise explanation of the task conducted and the study performed. All sensitive data about the participants were locked in a filing cabinet to which only the PhD supervisor and the research student had access. Participants were allocated a numeric identifier which was used to anonymise the data. The information linking participants to this numeric identifier was stored in a separate, locked cabinet.

2.2 CALCULATION OF SAMPLE SIZE

The choice of the sample size in an experimental study depends on the design of the experiment (the number of variables, the number of groups, and how conditions are compared) and on the type of statistical analysis that one intends to perform (parametric statistics, non-parametric, mixed models, etc). In this case, we were interested in the effect of two variables on one single group of participants. The two variables are *stress* and *position* (of the clusters) and they combine leading to 4 cross-conditions (cl1 str1 = the cluster is in the first syllable and the stress is on the first syllable; cl1 str2 = the cluster is in the first syllable and the stress is on the second syllable; cl2 str2 = the cluster is in the second syllable and the stress is on the second syllable cl2 str1 = the cluster is in the second syllable and the stress is on the first syllable). The sample size was calculated using an a-priori power analysis through the software G-power, a freely available source for sample size calculation. Since data were going to be analysed using a Two-Way ANOVA, a medium effect size for ANOVA was chosen as the aimed effect size. The chosen value is 0.06 η2 (chosen according to the guidelines of the Cognition and Brain Sciences Unit of the University of Cambridge http://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/effectSize). The software G-Power was then used to calculate the sample size needed to find an effect of this size (0.06 η2).[1] The value of the effect size was added in G-Power, specifying also the α error probability (0.05) and the power aimed for (0.8). These two values are the standard (conventional) choice for experimental research in psychology (Field, Miles & Field 2012). Conditions of the test were also defined: testing was going to take place in 1 group only and the conditions in the test were 4. The resulting sample size suggested by the software was 22 subjects.

2.3 PARTICIPANTS

The study was conducted on 22 Italian adults, 13 female, 8 male (1 not specified) aged between 20 years and 2 months and 24 years and 8 months, with a mean age of 21 years and 8 months, and a Standard Deviation of 1 year and 6 months, and no reported clinical conditions. All participants were Erasmus students, grew up as monolinguals, and started learning English from the age of 10 or 11, which is common in the Italian primary school system. All students had at least an intermediate command of

1 Since G-Power uses the f effect size (instead of η2) to compute this operation, a formula was used to obtain the f value starting from η2. The formula used is the following: $f = \sqrt{}\ \eta 2 / (1 - \eta 2)$. Using this formula, the resulting value of f is 0.26.

English (B1 in the European Framework). Before approaching the Erasmus students as potential participants, we considered testing pure monolinguals living in Italy, to avoid confounds with the L2. However, in the end we opted for Italian Erasmus students because, considering the current education system in Italy, in order for us to find pure monolinguals we would have needed to test elderly Italians and this would have created other problems in the interpretation of the results.

2.4 STIMULI

The task contained 40 pairs of nonwords. Half of the pairs consisted of identical nonwords and half of the pairs consisted of pairs of nonwords differing in one phoneme, generating a minimal pair. The nonwords used were trisyllabic and contained only the vowel /a/. Every nonword contained one phonological cluster of the type "plosive+liquid", for instance /t/ followed by /r/ as in /tra:kata/. This type of cluster was chosen because speakers never treat a sequence of plosive + liquid as belonging to two different syllables, but always as belonging to the onset of one syllable (Roach 2000). When nonwords in the pair differed in one phoneme, this phoneme was always part of the cluster, and the difference was always in one single trait, which was voicing. The reason for choosing this contrast could be defined as "historical", in the sense that this contrast has been used in a large number of studies on perception, dating back to the first investigations of categorical perception.[2] Clusters were positioned in the first or second syllable, and, similarly, stress could be positioned on the first or second syllable. For instance, pairs of differing nonwords included /tra:kata/ VS /dra:kata/ or /praka:ta/ VS /braka:ta/. The full list of stimuli is available in Appendix 1. The stimuli were recorded in the sound booth room of the School of Psychology and Clinical Language Sciences by a trained linguist whose first language was Italian. The software used was Audacity, running on a computer using Windows. The microphone used was an AKG D80, the mixer was a Behringer Mini Mon, the preamplifier was a B-tech phono-microphone preamplifier, and the speakers were Sony SS-CMD373. The length of the stimuli was controlled analogically, using the audio software Audacity. Once imported in Audacity, we ensured that the stimuli did not differ in length of more than 25ms (i.e. three percent of the total nonword length), using the Audacity timeline. The nonwords were recorded as "pairs", in the sense that each nonword recorded was followed in the recording by the other nonword forming a minimal pair. This corresponds to reading the list of nonwords in Appendix 1 row by row.

2.5 PROCEDURE

The testing session lasted approximately 30 minutes. Participants were asked to read the information sheet and sign the consent form. After giving consent, participants were asked to sit comfortably in front of the computer and the researcher gave the instructions for the task orally. The researcher then left the room and let the participant read the instructions on the screen and put on the headphones. The instructions on the screen guided the participant through the testing session. The task was programmed on the software E-prime. Participants were presented with pairs of

2 For a review about voicing studies, see Hoonhorst et al. 2011.

nonwords. Each presentation of a pair consisted of a rapid succession of two nonwords. The nonwords in each pair could either be identical or form a minimal pair, differing in one phoneme. A white sticker was placed over the letter "w" on the keyboard, and a black sticker was placed over the letter "b" on the keyboard. Participants were asked to press white when they thought the two nonwords presented as part of a pair were identical and to press black when they thought that the two nonwords were different. The exact text of the instructions was: "In the experiment you will hear pairs of invented words presented in rapid succession, one next to the other. Sometimes the two words will be identical, sometimes they will be slightly different. Press white when you think the two words are identical, press black when you think that they are slightly different." Considering that the position of the cluster was manipulated and considering that the position of stress was manipulated, there were four possible conditions in which the two nonwords differed, and four possible conditions in which the two nonwords were the same. The combination of possible stimuli is presented in Table 1. In the table, stress is represented in its main manifestation in Italian: the lengthening of vowels. Stressed syllables, thus, are syllables that in this table end with the symbol ":", representing lengthening. As Krämer points out (2009), lengthening in penultimate and antepenultimate syllables in Italian bears different properties, and is more pronounced in the former. This variation was accounted for in this experiment by the fact that stimuli were recorded by a native speaker of Italian, and they were uttered to sound "as natural as possible". Stimuli were presented randomly, using the random function of E-prime. The order was different for every subject. Each minimal pair was never repeated twice in one experimental session.

Description of stimuli	**Prime**	**Different**	**Same**
Cluster in the first syllable, stress on the first syllable	/tra:kata/	/dra:kata/	/tra:kata/
Cluster in the first syllable, stress on the second syllable	/traka:ta/	/draka:ta/	/traka:ta/
Cluster in the second syllable, stress on the second syllable	/katra:ta/	/kadra:ta/	/katra:ta/
Cluster in the second syllable, stress on the first syllable	/ka:trata/	/ka:drata/	/ka:trata/

TABLE 1. Example of stimuli. The four conditions are presented together with their contrasting items.

Each trial (i.e. each presentation of a pair) was composed of four different slides: the first slide, named "fixation", introduced the subject to the new trial. It was one second long, it did not contain any sound and it only consisted of a visualization of the symbol # in the centre of the screen. This slide was included to catch the attention of the participant and make them focus on the task. The second slide, named "soundout1", consisted of the presentation of the first sound in the pair, which in Table 1 is referred to as "prime". During the presentation of this slide, the screen was completely black (the stimulus was presented aurally). The length of this slide was 1000 ms, although the length of the stimuli was, on average, 730 ms (SD 39 ms). The length was also calculated post-hoc with the automatic measures of the Voice Activity Detector (VAD) software and the automatic text-grid of Praat.[3] The first sound was then followed by the target sound, presented again in a black slide (soundout2). The final slide was named "TextDysplay", and it contained an arrow (→) that indicated the end of the trial and the approach of a new trial. The final slide had a length of 1s (1000 ms).

2.6 SCORING

The measures collected by E-prime were the following: the answer the subject gave (whether white, black, a non-valid key or no answer) and the time they took to provide an answer. Subjects had a limit of 1000ms to give an answer (from the beginning of the target stimulus). If no answer was given within 1000 ms, the software coded "no answer". Accuracy was calculated by dividing the number of errors by the number of given answers. Since the task was measuring the ability to discriminate phonological contrasts, we only used the trials in which the two elements contrasted in the pair were different. The pairs in which the two elements were identical were used as fillers. Given the type of coding used, accuracy could be any number between 0 (none of the given answers is wrong) and 1 (all of the given answers are wrong). For instance, if, in a given condition, the subject answered 5 times and none of the answers was wrong, the accuracy for that participant in that condition was coded as 0, since this is the result of the ratio 0/5. If the subject answered 5 times and 4 of them were wrong, accuracy was coded for that participant in that condition as 0.8, since this is the result of the ratio 4/5. If, for instance, the subject answered only 2 times, but both answers were wrong, accuracy was coded as 1, since this is the result of the ratio 2/2. A similar choice was made in previous research (i.e. Mora 2005).

3 RESULTS

Condition	Minimum	Maximum	Mean	Std. Error	Std. Dev.
cl1 str1	0	1	.396	.050	.231
cl1 str2	0	.80	.379	.058	.265
cl2 str2	0	.80	.249	.054	.248
cl2 str1	0	1	.569	.062	.286

TABLE 2. Descriptive statistics, including min and max values for accuracy in the four conditions, as well as mean, standard deviation and standard error.
Key: cl1 str1 = the cluster is in the first syllable and the stress is in the first syllable; cl1 str2 = the cluster is in the first syllable and the stress is on the second syllable; cl2 str2 = the cluster is in the second syllable and the stress is on the second syllable; cl2 str1 = the cluster is in the second syllable and the stress is on the first syllable; accuracy is presented as the proportion of correct responses.

3 The assistance of Professor Radek Skarnitzl of Charles University in Prague is gratefully acknowledged. Note that the automatic and analogical measures of length substantially overlap. In fact, the analogical measure may actually be more accurate, since the software is unable to interpret final vowel lengthening and is relatively inaccurate in the detection of the beginnings of words starting with voiced plosives.

In order to ensure that participants were engaging in the task and not answering randomly, D-prime values were calculated. D-prime is a statistical procedure that offers an estimate of a participant's bias toward one answer or the other in the same/a different task. This estimate indirectly shows the quality of engagement in the task. If, for example, in our experiment one of our participants was consistently answering "different" to every single trial (independently of whether the nonwords were same or different), their accuracy would be at 100%, but this measure would be totally misleading, since the participant is not really succeeding in the discrimination of any of the items, and in fact is not engaging in the task. Accuracy in one condition is only meaningful if we can demonstrate that participants are not biased in one direction or the other, but in each individual trial they are open to both possible answers (*same* and *different*). D-prime values give an estimate of this bias, and they were calculated for each participant. Using one sample t-tests, D-prime values were then compared to 0 and to 1.[4] D-prime values showed to be significantly bigger than 0, t (20) = 6.51, $p < .0001$, and significantly bigger than 1, t (20) = 3.92, $p = .0008$ two tailed. These results suggest that participants were not answering the discrimination task randomly, and it also indicates that performance was better than chance and that it was accurate in more than 70% of cases (Macmillan & Creelman 2005).

Once ascertained that participants were engaging in the task, we proceeded with an analysis of their responses. Since the study compares means in 4 different conditions (generated by two variables) in 1 group of people, a repeated measures ANOVA was conducted (2×2). Factor one was cluster position (the cluster could be either in the first or in the second last syllable), factor two was stress (the cluster could be either stressed or not). Descriptive statistics are presented in Table 2. The ANOVA shows a significant effect of stress, F (20,1) = 8.42, $p = .009$, no significant effect of cluster position, F (20,1) = .17, $p > .05$, and a significant interaction, F (20,1) = 30.08, $p < .001$. Figure 1 shows visually the comparison of means and the factors used in the ANOVA.

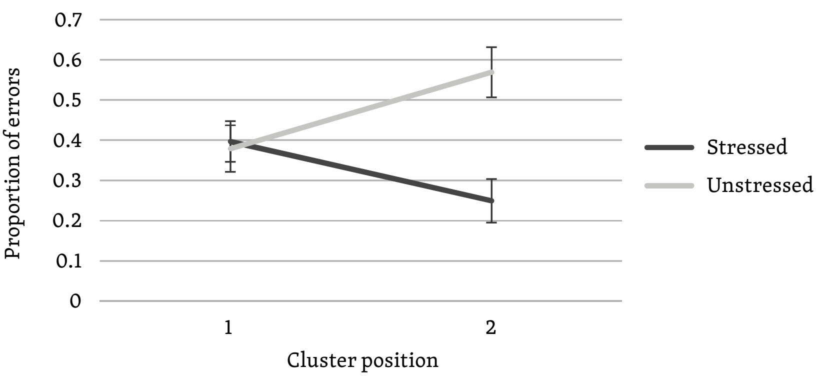

FIGURE 1. Errors across conditions

4 A thorough explanation of the rationale behind this procedure can be found on the UCLA phonetics website: http://phonetics.linguistics.ucla.edu/facilities/statistics/dprime.htm

Post-hoc t-tests were carried out to investigate which contrasts led to the significant stress effect and to the interaction. Six contrasts were analyzed. For this reason, the Bonferroni adjusted alpha used is p = .05/6 = 0.008. This approach is rather conservative and reduces significantly the probability of finding a false positive. The results obtained are summarized in Table 3.

Contrast	***t* value**	***p* value**
Cl1Str1 — Cl1Str2	t (20) = .275	p = .78
Cl2Str2 — Cl2Str1	t (20) = −5.58	p < .001 *
Cl1Str1 — Cl2Str2	t (20), 2.87	p = .009
Cl1Str2 — Cl2Str1	t (20), −2.88	p = .009
Cl1Str1 — Cl2Str1	t (20) = −2.48	p = .02
Cl1Str2 — Cl2Str2	t (20) = 1.72	p = .09

TABLE 3. Post-hoc analyses
Key: cl1 str1 = the cluster is in the first syllable and the stress is on the first syllable; cl1 str2 = the cluster is in the first syllable and the stress is on the second syllable; cl2 str2 = the cluster is in the second syllable and the stress is on the second syllable; cl2 str1 = the cluster is in the second syllable and the stress is on the first syllable. * indicates a significant result.

According to the post-hoc analysis, the contrast that is driving the ANOVA is the contrast between stressed and unstressed syllables when the cluster is in medial position. This suggests that the significance of the stress effect is uniquely driven by the difference between stressed and unstressed syllables in medial position, and it suggests that the interaction is driven by the fact that the contrast is significant in medial position but absent in nonword initial position.

4 DISCUSSION

The aim of this experiment was to investigate whether word initial positions are salient in perception in Italian, and how (if) stress interacts with their saliency. The results showed that position saliency does interact with stress: a stress effect was consistently found only when clusters were in word medial syllables. In word initial positions, instead, stressed and unstressed syllables were discriminated with similar accuracy.

As mentioned in the Introduction, stress patterns are of pivotal importance during language acquisition. When children have a sufficient number of words in their vocabulary, they can extract the stress pattern of these words and try to use this pattern as a cue for further speech segmentation (Echols 1996). For instance, English children may assume that words start with a stressed syllable, while French children may assume that they finish with a stressed syllable, since it is very likely that the first words they learn will follow this pattern. Thus, stress is a fundamental cue in early speech segmentation and it remains a crucial property for adults, since the stress pattern of a language reveals which positions are salient in continuous speech (Ambridge & Lieven 2011).

Our results suggest that in tasks using nonwords in isolation (such as the task used in this experiment) stress saliency may interact with a cross-linguistic general principle, that of word beginning saliency, leading to the presence of stress effects only in syllables that are non-initial. The reason behind the deactivation of a stress effect in initial syllables is not clear, but a formal explanation may be attempted. In the next section we offer two possible explanations for this result.

The first proposed explanation is that of a ranking of constraints: in Beckman (1998), the word beginning saliency principle is described as a phonological constraint, rather than a phonological rule. Constraints are different from rules in that they can be violated (Prince & Smolensky 1997): they are not inevitable operations but rather tendencies that guide and influence our parsing of speech (as well as our generation of speech, though production is not the focus of this paper). Constraints, differently from rules, can contradict one another. Given the existence of these contradictions, it is assumed that speakers will not always comply to a constraint. For example, the statement that there is a constraint establishing that word beginnings are salient does not imply that every speaker in every context will concentrate their resources on word beginnings, but rather that this will be a consistent tendency, which will take place as long as no other more powerful constraint "forbids" it. The different power or importance of constraints is sometimes described as a hierarchy. The word beginning constraint is, according to Beckman, highly ranked (Beckman 1998). This means that most of the other phonological constraints will not be able to lead to its violation, and only constraints positioned in a higher position will lead to its violation. Our data suggest that a hierarchical block of this type may be taking place with stressed syllables. If we assume the existence of a stress constraint, the interaction between word beginning effects and stress becomes predictable using ranking. Specifically, if we describe stress saliency as a stress constraint that establishes that stressed positions are more salient than unstressed positions, and we assume this constraint to be ranked in a lower position than Beckman's constraint, then we would predict the word beginning saliency effect to inhibit the stress effect. In other words, the stress effect would become visible only in non-initial positions. This is indeed what we report here for Italian. The proposal for the existence of these two constraints is consistent with psycholinguistic data. From a phonetic and psycholinguistic point of view, it is well acknowledged that stressed positions are salient and less prone to disruption (Roach 2000), and, on the other hand, there is evidence from lexical access studies showing that word beginnings are salient and more important in selecting the target word (Zwitserlood 1989). However, there are two problems with this explanation: first, this proposal does predict a main effect of position to appear, but we do not observe a main effect of position directly (we only observe that the stress effect is dependent on position). Second, if the position constraint outranks the stress constraint, we would expect initial syllables to be overall very accurate, whether they are stressed or not. We find, instead, that performance is at chance in initial syllables (whether they are stressed or not).

A second possible explanation of this result is a perceptual one. In Italian, stress is cued by vowel lengthening and not by intensity (Krämer 2009). For this reason, when stress is in word initial position, participants have no way of knowing whether the syllable is going to be stressed or not, when they hear the cluster. The cluster was always placed in the onset of the syllable, so participants had to wait until after the cluster was uttered to understand whether the syllable was stressed. This is not the case for word medial clusters, since participants could quickly realize that if stress was not appearing in word initial position, it would indeed appear in word medial position. In other words, when stress was in medial position, participants could anticipate its placement before actually hearing it. This interpretation predicts an *intermediate* result with initial clusters (and no effect of stress), and a clear effect of stress when clusters are in word medial position, which is indeed what we obtained. Such an interpretation corroborates the suggestion we outlined in the introduction: the word beginning saliency principle observed in English may be driven by stress being often on the first syllable in that language, and thus we do not expect the same word beginning principle found in English to occur also in Italian. As a corollary to this explanation, we add that the performance with medial clusters might have been influenced by how prototypical a nonword is (and not just by the acoustic saliency cued by stress). As discussed in the introduction, Italian words tend to have stress on the penultimate syllable. In this experiment, the condition where clusters were in a stressed medial syllable was the most accurate of all. If this interpretation of the data is correct, accuracy in this condition may have been influenced by the fact that nonwords with medial stress are the closest to the Italian prototype; on the other hand, nonwords with an unstressed medial cluster are less expected and thus processed with lower accuracy.

In short, while the overall intermediate performance with initial clusters may be explained purely on the basis of a perceptual phenomenon, the pattern observed in the medial clusters may have been influenced by how prototypical the syllable was (and not just by the acoustic saliency of stress). Further experiments could disentangle pure effects of stress from prototype effects, to ascertain the role of both in inducing saliency.

5 CONCLUDING REMARKS

The present study investigated, by psycholinguistic means, the interaction between word beginning and stress saliency, and it showed that stress operates in interaction with the word beginning principle: in Italian adults, according to our data, stress saliency effects are large in medial syllables but inhibited in word initial positions. This pattern may be due to a ranking of phonological constraints, where word position outranks stress in terms of importance, or it may be due to a combination of perceptual phenomena and "prototype effects".

OPEN ACCESS

REFERENCES

Ambridge, B. and E. V. Lieven (2011) *Child language acquisition: Contrasting theoretical approaches*. Cambridge: Cambridge University Press.

Beckman, J. N. (1998) *Positional faithfulness*. Amherst: University of Massachusetts PhD dissertation.

Beckman, J. N. (2013) *Positional faithfulness: an optimality theoretic treatment of phonological asymmetries*. London: Routledge.

Benasich, A. A. and P. Tallal (2002) Infant discrimination of rapid auditory cues predicts later language impairment. *Behavioural brain research* 136, 31–49.

Browman, C. P. (1978) Tip of the tongue and slip of the ear: implications for language processing. *University of California Working Papers in Phonology* 42, 1–159.

Dehaene, S. (2009) *Reading in the brain, the new science of how we read*. New York: Penguin Books.

Domahs, U., I. Plag and R. Carroll (2014) Word stress assignment in German, English and Dutch: quantity-sensitivity and extrametricality revisited. *The Journal of Comparative Germanic Linguistics* 17/1, 59–96.

Echols, C. H. (1996) A role for stress in early speech segmentation. In: Morgan, J. L. and K. Demuth (eds) *Signal to Syntax*, 151–170. New York: Psychology Press.

Field, A., J. Miles and Z. Field (2012) *Discovering statistics using R*. New York: Sage Publications.

Guasti, M. T. (2004) *Language acquisition: the growth of grammar*. Cambridge (MA): The MIT Press.

Guttorm, T. K., P. H. Leppänen, J. A. Hämäläinen, K. M. Eklund and H. J. Lyytinen (2010) Newborn event-related potentials predict poorer pre-reading skills in children at risk for dyslexia. *Journal of Learning Disabilities* 43/5, 391–401.

Harris, J. (2011) Deletion. In: van Oostendorp, M., C. J. Ewen, E. V. Hume and K. Rice (eds) *The Blackwell Companion to Phonology*, 1597–1621. Oxford: Wiley-Blackwell.

Hayes, B. (1982) Extrametricality and English stress. *Linguistic Inquiry* 13, 227–276.

Hayes, B. (2012) *How predictable is Italian word stress*. Hsinchu: National Chiao Tung University.

Hoonhorst, I., V. Medina, C. Colin, E. Markessis, M. Radeau, P. Deltenre and W. Serniclaes (2011) Categorical perception of voicing, colors and facial expressions: a developmental study. *Speech Communication* 53/3, 417–430.

Jusczyk, P. W., A. Cutler and N. J. Redanz (1993) Infants' preference for the predominant stress patterns of English words. *Child development* 64/3, 675–687.

Krämer, M. (2009) *The phonology of Italian*. Oxford: Oxford University Press.

Macmillan, N. A. and C. D. Creelman (2005) *Detection Theory: A User's Guide*. New York: Lawrence Erlbaum Associates.

Männel, C. and A. Friederici (2008) Event-related brain potentials as a window to children's language processing: from syllables to sentences. In: Sekerina, I. A., E. M. Fernández and H. Clahsen (eds) *Language acquisition and language disorders: Vol. 44. Developmental psycholinguistics: On-line methods in children's language processing*, 29–72. Amsterdam: John Benjamins.

Marshall, C. R. and H. K. van der Lely (2009) Effects of word position and stress on onset cluster production: Evidence from typical development, specific language impairment, and dyslexia. *Language* 85/1, 39–57.

Marslen-Wilson, W. and P. Zwitserlood (1989) Accessing spoken words: the importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance* 15/3, 576–585.

Mehler, J., E. Dupoux, T. Nazzi and G. Dehaene-Lambertz (1996) Coping with linguistic diversity: the infant's viewpoint. In: Morgan, J. L. and K. Demuth (eds) *Signal to Syntax*, 101–116. New York: Psychology Press.

Mora, J. C. (2005) Lexical knowledge effects on the discrimination of non-native phonemic contrasts in words and nonwords by Spanish/

Catalan bilingual learners of English. In: Hazan, V. and P. Iverson (eds) *ISCA Workshop on Plasticity in Speech Perception, Senate House, London, UK, June 15–17, 2005*.
Nooteboom, S. G. (1981) Lexical retrieval from fragments of spoken words: Beginnings vs endings. *Journal of Phonetics* 9/4, 407–424.
Pitt, M. A. and A. G. Samuel (1995) Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology* 29/2, 149–188.
Prince, A. and P. Smolensky (1997) Optimality: From neural networks to universal grammar. *Science* 275/5306, 1604–1610.
Roach, P. (2000) *English Phonetics and Phonology*. Cambridge: Cambridge University Press.
Saffran, J. R., R. N. Aslin and E. L. Newport (1996) Statistical learning by 8-month-old infants. *Science* 274/5294, 1926–1928.
Smith, J. L. (2004) *Phonological augmentation in prominent positions*. New York: Routledge.
Walley, A. C. (1988) Spoken word recognition by young children and adults. *Cognitive development* 3/2, 137–165.
Zwitserlood, P. (1989) The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition* 32/1, 25–64.

Luca Cilibrasi
Faculty of Arts, Charles University
nám. Jana Palacha 2, 11638 Praha 1
ORCID ID: 0000-0001-8262-0542
luca.cilibrasi@ff.cuni.cz

Vesna Stojanovik
School of Psychology and Clinical Language Sciences, University of Reading
Harry Pitt Building Earley Gate, Reading, RG6 7BE
ORCID ID: 0000-0001-6791-9968
v.stojanovik@reading.ac.uk

OPEN ACCESS

APPENDIX

	Unvoiced	Voiced
Cl1 str1	tra:kata pla:kata pra:kata kla:kata kra:kata	dra:kata bla:kata bra:kata gla:kata gra:kata
Cl1 str2	traka:ta plaka:ta praka:ta klaka:ta kraka:ta	draka:ta blaka:ta braka:ta glaka:ta graka:ta
Cl2 str2	katra:ta kapla:ta kapra:ta kakla:ta kakra:ta	kadra:ta kabla:ta kabra:ta kagla:ta kagra:ta
Cl2 str1	ka:trata ka:plata ka:prata ka:klata ka:krata	ka:drata ka:blata ka:brata ka:glata ka:grata

Complete list of stimuli