



# When two newly-acquired words are one: New words differing in stress alone are not automatically represented differently

Simone Sulpizio<sup>1</sup>, James M. McQueen<sup>2,3,4</sup>

<sup>1</sup> Department of Cognitive Science and Education, University of Trento, Italy

<sup>2</sup> Donders Institute for Brain, Cognition, and Behaviour, Centre for Cognition,  
Radboud University Nijmegen, Nijmegen, The Netherlands

<sup>3</sup> Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, The Netherlands

<sup>4</sup> Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

simone.sulpizio@unitn.it; j.mcqueen@pwo.ru.nl

## Abstract

Do listeners use lexical stress at an early stage in word learning? Artificial-lexicon studies have shown that listeners can learn new spoken words easily. These studies used non-words differing in consonants and/or vowels, but not differing only in stress. If listeners use stress information in word learning, they should be able to learn new words that differ only in stress (e.g., *Blulo-biNulo*). We investigated this issue here. When learning new words, Italian listeners relied on segmental information; they did not take stress information into account. Newly-acquired words differing in stress alone are not automatically represented as different word forms.

**Index terms:** lexical stress, spoken word recognition, Italian, learning new words.

## 1. Introduction

Talking is one of the most common activities in everyday life. To understand someone speaking to us, we have to convert the acoustic signal produced by that speaker into a meaningful message. To achieve this, listeners perform word recognition: They access the candidate words that match the incoming phonetic information and then recognize the best matching candidate. During this process, listeners use all the incoming information in the acoustic signal [1; 2]. Recent studies in Dutch and Italian have highlighted that, when recognizing words, listeners exploit not only the segmental information in the signal (i.e., the information specifying individual vowels and consonants), but also the available suprasegmental information (e.g., information specifying the word's lexical stress pattern [3; 4]). Listeners appear to use both types of information as soon as it comes available to them during the temporal unfolding of a spoken word.

In everyday listening, however, people are not always able to match the acoustic information in the speech signal with an existing lexical entry. We can hear a word that we do not know. In this case, we have to learn the new word. We have to learn what the word means, and what it sounds like. That is, we have to establish a new lexical entry for a previously unknown phonetic signal. During this learning process, we create a new lexical representation that we will be able to use for speech comprehension when we hear that particular acoustic-phonetic signal again. To do so, however, we also have to link the new lexical representation with a referent (e.g., an object or an event).

An important question about word-learning concerns the information used in building new lexical representations. Do segmental and suprasegmental information both play a role in the earliest stages of word learning? We ask here whether the stress pattern of a newly-acquired word automatically becomes part of its emerging lexical representation.

On the one hand, if we consider languages such as Italian, Dutch and English, in which stress position is not fixed – that is, languages in which stress does not appear always in the same place in the word (as, e.g., in French, Finnish or Slovak) – then one might expect that stress information would be relevant even at the earliest stages of lexical learning. In these languages, lexical stress information alone can distinguish between words. Furthermore, as we have just mentioned, listeners use stress information, alongside segmental information, to recognize well-known words as soon as that information is available to them [3; 4]. We might then hypothesize that both segmental and suprasegmental information will drive lexical learning.

On the other hand, when listeners are learning new words, they might weigh segmental and suprasegmental information differently. They could choose to focus their attention on the segmental characteristics of a new word first, perhaps because prior experience with their native language has taught them that the segmental properties of words tend to be more informative than their suprasegmental properties (e.g., a random pair of words is much more likely to differ segmentally than suprasegmentally). Thus, while listeners learn quite easily to associate new pairs of words to new objects when their names are segmentally different [5], they could have serious difficulties performing the same task when the new pairs of words are segmentally identical and differ only in stress pattern. Consider a variable-stress language such as Italian. Although stress minimal pairs exist (e.g., *ANcOra* ‘anchor’ vs. *anCOra* ‘again’; capital letters indicate the stressed syllable), words usually differ from each other at the segmental level. The fact that the greatest source of diversity between words is at the consonant and vowel level [6] might thus affect the way in which Italians learn new words. During their early exposures to novel words, Italian listeners might establish their first lexical representations based on consonant and vowel information (i.e., by focusing on acoustic-phonetic segmental cues). In this first stage, they might ignore the prosodic cues related to lexical stress, knowing that very few Italian words differ only in stress pattern.

A good way to test these alternative hypotheses is to use an artificial lexicon composed of non-words and unfamiliar objects. Participants hear novel words and learn to associate them to novel objects. Most of the studies conducted with an artificial-lexicon paradigm have used stimuli with segmental differences [5; 7; 8]. In all these studies, participants learned fairly well to associate the non-words to the novel objects. But this may not necessarily be the case if the non-words differ from each other only at the suprasegmental level, that is, if they differ only in their stress pattern (e.g., *Blulo* and *biNulo*). This was the situation that we tested here.

Italian three-syllable words bear stress mainly on the penultimate syllable (e.g., *maTta*, ‘pencil’; 80% of words

have this stress pattern [9]). Most other words (18%) have antepenultimate stress (e.g., *Tavolo*, ‘table’). Only very few words (2%) have stress on the last syllable (e.g., *colibri*, ‘hummingbird’); we do not consider them further. Furthermore, stress position in Italian trisyllabic words is not predicted by rule [10]. In addition, there is no vowel reduction to schwa in Italian unstressed syllables. For these reasons, Italian is a good candidate to test whether the stress patterns of newly-acquired words are automatically built into their emerging lexical representations. We can control the materials such that they differ only in stress pattern, and we have a situation where, in spite of the bias favoring penultimate stress, the stress pattern of a new word is not fully predictable, and thus is something that, at least in principle, has to be stored as part of each lexical representation (i.e., if stress information is stored at all, it has to be stored in the representations of the individual new words).

We used a set of non-words that we recorded twice. In this way, we obtained minimal stress pairs: one non-word of the pair had penultimate stress (e.g., *biNUlo*); the other one had antepenultimate stress (e.g., *BInuLo*). During a training phase, participants learned to associate the novel words to novel objects, performing a two- or four-alternative forced-choice task. They heard the non-words at the end of a carrier sentence (e.g., *Clicca sul BInuLo*, ‘Click on the BInuLo’) and they had to select the corresponding non-object on a computer screen (choosing between either 2 or 4 possibilities). During the training phase, we never showed both elements of a given minimal pair on the same screen. Then, during the subsequent test phase, participants performed the same task, but we showed both members of a minimal pair among the alternatives displayed on the screen.

In this way we could test whether listeners were able to recognize the target non-object (e.g., the object associated with *BInuLo*) and discard its stress competitor (e.g., the object associated with *biNUlo*). If this were the case, then it would show that Italians can exploit all the acoustic information in the signal during the earliest stages of word learning. In contrast, if Italian listeners do not necessarily pay attention to lexical stress and do so only when the learning situation explicitly requires it, then they will have trouble in distinguishing the target stimulus (e.g., *BInuLo*) from its segmentally identical competitor (e.g., *biNUlo*). This is because, during training, it was never necessary to use stress information to learn the name-object associations: *BInuLo* and *biNUlo* could each separately be identified and learned on the basis of segmental information alone.

## 2. Method

### 2.1. Participants

Thirty-one students (mean age: 24.6, sd: 5.2) from the University of Trento took part in the experiment. They received course credit for their participation. All participants were Italian native speakers with no known hearing problems and with normal or corrected-to-normal vision.

### 2.2. Materials

Sixteen three-syllabic non-words were created as stimuli (*berzifo*, *binulo*, *botalo*, *canvilo*, *confuro*, *curfino*, *desico*, *dunzico*, *gavilo*, *goliso*, *patuco*, *pencilo*, *pindumo*, *tefubo*, *tolaco*, and *tudero*). Each non-word was recorded twice, once using penultimate stress (e.g., *biNUlo*), and once using antepenultimate stress (e.g., *BInuLo*). In this way, 16 critical pairs were obtained in which the two non-words were segmentally identical and differed only in stress position.

Thirty-two line drawings of nonsense objects were randomly selected from a database of non-objects (Non-existing Objects Database, [www-server.mpi.nl/experiment-pictures/production-pictures/](http://www-server.mpi.nl/experiment-pictures/production-pictures/)). The non-objects were randomly assigned to the non-words.

A female native Italian speaker, naïve about the experiment’s purpose, recorded the stimuli in a sound-attenuated room (sampling at 44 kHz, 16 bit resolution, mono). Stimuli were recorded at the end of the target sentence “*Clicca sul*” (‘Click on the’). Each stimulus was also recorded at the end of a feedback sentence (e.g., *Ora puoi vedere di nuovo il biNUlo*, ‘Now you can see the biNUlo again’). We wanted to establish whether Italians could learn about specific acoustic cues to stress. Therefore, using PRAAT [11], we controlled the materials by neutralizing the amplitude and duration of the first two vowels. That is, for each pair of non-words, we replaced the original amplitude and duration values of each vowel with the average of those values across the pair (e.g., for the pair of /i/ tokens in *BInuLo* and *biNUlo*; for further details on acoustic measures and procedure, see [4]). Stimuli were spliced back into the carrier sentences.

### 2.3. Procedure

The experiment was composed of two phases: a training phase and a test phase. During the training phase, participants learnt to associate non-words to non-objects. During the test phase, they heard the non-words and they had to recognize the corresponding non-objects. During the test phase we recorded participants’ eye-movements using a head-mounted Eyelink II System (sampling rate: 500 Hz).

The training phase was composed of 3 blocks. Each stimulus was presented 3 times within each block, for a total of 96 trials in each training block (32 non-words x 3 repetitions). In Blocks 1 and 2, participants had to select the target non-object from two alternatives that belonged to different critical pairs (e.g., we displayed *BInuLo* and *toLaco*). In contrast, in Block 3 we displayed four non-objects; again, the non-objects belonged to different critical pairs (e.g., *BInuLo*, *toLaco*, *PENcilo*, and *canVilo*). Thus, during the training phase, listeners never saw on the same screen both the non-objects whose names formed a critical minimal pair. In this way, participants were not required to focus explicitly on the stress difference.

The procedure for the training phase was as follows. Each trial started with a fixation cross in the center of the screen, displayed for 500 ms. Then the non-objects appeared on the screen and remained there until participants clicked the mouse button. The auditory instruction (carrier sentence plus target word, e.g., *Clicca sul BInuLo*) was played over headphones. Participants had to click the mouse on the target non-object whose name was heard at the end of the carrier sentence. At the same time, a sentence was played to indicate if the response was correct (*giusto*, ‘right’) or not (*sbagliato*, ‘wrong’). Then the target non-object was displayed again in the center of the screen and the feedback sentence (e.g., *Ora puoi vedere di nuovo il BInuLo*) was played.

Before the test, the eye-tracker was mounted and calibrated. The test phase had one block and each target was shown twice. This phase directly followed the training phase. On each trial, participants heard a target non-word at the end of the carrier sentence and they had to select the corresponding non-object among 4 possible alternatives displayed on the screen. The four possible choices belonged to two critical minimal pairs (e.g., *BInuLo* and *biNUlo*; *TOLaco* and *toLaco*; see Figure 1). Stimulus order was randomized within the block, as was target position on the screen.

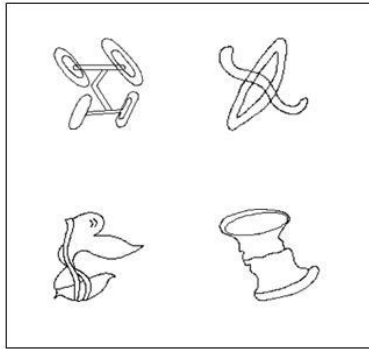


Figure 1: Example of a 4-alternative display (used in Block 3 in the training phase and in the test phase)

In the test phase, each trial was structured as follows. First, a fixation cross was displayed, centered on the screen, for 500 ms. Then four non-objects appeared on the screen and remained there either until participants clicked the mouse button or for a maximum of 5000 ms. A white screen was used in the inter-stimulus interval of 480 ms. The four non-objects were centered in the four quadrants of the screen. The auditory instructions were again played over headphones. Participants had to click the mouse on the target non-object whose name was heard at the end of the carrier sentence. There was no feedback during the test phase.

### 3. Results

We compared the percentage of correct and incorrect responses in the training phase and in the test phase. In this way, we were able to evaluate two different things: first, whether listeners had learnt the non-objects' names by the end of the training phase; second, whether listeners had learnt the stress information associated with each non-object's name, as measured in the test phase.

During the training phase, participants learned to associate the non-words to the non-objects: They were able to distinguish the targets from their distractors. The percentage of correct responses increased from 65% in the first training block to 84% of the last training block. Moreover, in the third block – where there were four alternatives – all the incorrect alternatives were selected equally often (see Table 1).

Table 1: Percentage of responses for each condition in the three blocks of the training phase

Blocks	Alternatives			
	Target	Choice 2	Choice 3	Choice 4
Block 1	65%	35%	-	-
Block 2	85%	15%	-	-
Block 3	84%	4%	6%	6%

It is important to remember, however, that even though the listeners appeared to be able to learn to select the target non-objects accurately from among four possible choices, the possible alternatives displayed on the screen during the final training block differed amongst each other at the segmental level. In contrast, during the test phase, participants saw on the screen non-objects that did not differ segmentally. They saw a target non-object (e.g., *Bmulo*), its stress competitor (e.g., *biNUlo*), and two segmentally unrelated distractors that themselves differed only in stress (e.g., *toLAco* and *TOLaco*).

The pattern of results obtained in the test phase was quite different from that obtained in the training phase. The

percentage of correct responses dropped dramatically from 84% in the third training block to 54% in the test phase. In the overwhelming majority of the incorrect responses, listeners clicked on the stress competitor (42%) instead of one of the distractors (or indeed the target; see Figure 2). Eye-tracking data show the same result: listeners were not able to distinguish between target and competitor (see Figure 3).

We performed two different statistical comparisons to test what listeners had learnt during the training phase. In the first analysis, we compared the responses given in the test phase to the critical minimal pairs (the target and its stress competitor) with the responses given to the distractor pairs. This provided a measure of whether listeners could discriminate among stimuli that differed at the segmental level. The comparison between the critical and distractor pairs showed that the responses given in these two categories significantly differ ( $\chi^2 = 51.36$ ,  $df = 1$ ,  $p < .01$ ): During the training phase, listeners had clearly learnt the segmental differences in the stimuli.

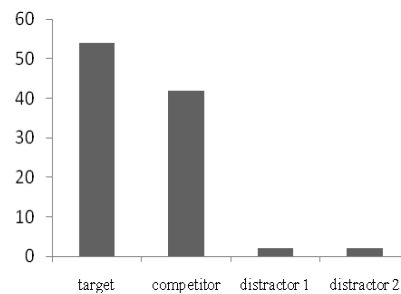


Figure 2: Percentage of responses for each condition in the test phase

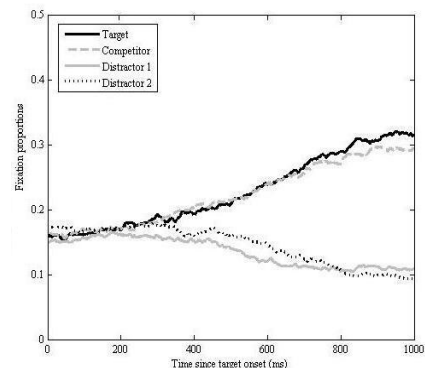


Figure 3: Fixation proportions to targets, competitors, and distractors over time.

In the second analysis, we compared the responses given to the targets with the responses given to their competitors. This comparison allowed us to test whether listeners were able to discriminate between the target (e.g., *Bmulo*) and its stress competitor (e.g., *biNUlo*). The comparison showed that the responses given to the targets did not differ from the responses given to their competitors ( $\chi^2 = 0.5$ ,  $df = 1$ ). By subjects and by items t-test comparisons of target and competitor fixation proportions showed the same result (both  $t_s < 1$ ). It appears that if two non-words differed only in their stress patterns, then the listeners were not able to identify the target non-word and discard its segmentally identical competitor.

### 4. Discussion

We tested whether, in Italian, lexical learning focuses more on segmental material than on suprasegmental material, or, alternatively, whether stress information is automatically used alongside segmental information in word learning. At least at

the beginning of the learning process, it appears that Italian listeners take consonant and vowel differences into account, but do not consider stress differences.

At the end of the training phase, listeners performed well on the recognition task: they selected the right non-object with high accuracy. But their accuracy dropped to chance level in the test phase: They were not able to distinguish target non-objects (e.g., *BInulo*) from their stress competitors (e.g., *biNUlo*). In line with previous research [5; 7], when new words segmentally differed from each other (e.g., *BInulo* and *toLAcO*), listeners recognized the non-objects fairly well after relative few exposures (i.e., at the end of the training phase). But when possible candidates differed only in stress pattern (i.e., in the test phase), listeners were not able to distinguish between the two alternatives.

The present findings show that when Italian listeners are learning new words and are not encouraged to focus on suprasegmental information, they focus instead solely on segmental information. Italians may build their initial lexical representations using consonant and vowel information because it is the main source of word diversification [6; 12] and because it is sufficient to establish a lexical representation using segmental information alone. Lexical stress information may be built into lexical representations later.

These findings are consistent with those showing that stress information did not necessarily play a substantial role in word learning in an artificial-lexicon study with English adults [8], where listeners used mainly segmental information to recognize words. Curtin [13], however, found different results when investigating the early stages of word learning in English infants. In her experiment, 12-month-olds learnt to associate two non-words with two new objects; the two non-words were segmentally identical, but differed in stress. Then, during the test phase, infants saw on the screen an object and they heard either its name or the name of the other object, which had the different stress pattern. Measuring looking time, Curtin found that infants learnt to associate the two novel words with the two novel objects, even if their names differed only in stress pattern (for similar results, see [14]).

A resolution for the apparent contradiction between these findings [13] and our own lies in an important procedural difference. The infant experiments involved only two items. This could encourage participants to focus on the stress difference. That is, the infant listeners could identify that the two stimuli differed only in stress pattern and hence that they had to pay attention to the suprasegmental information. A similar conclusion follows from Sulpizio and McQueen's study with adult listeners [4]. In a study on Italian lexical stress, they used the same materials and almost the same procedure as those used here. The critical difference was that, during training, participants saw both elements of the minimal stress pairs (e.g., *BInulo-biNUlo*) on the same screen. Participants were thus forced to attend to stress information and hence encouraged to build stress into their nascent lexical representations. The results showed that the participants did indeed do so. Note that these findings show that Italians could learn about the stress patterns of the present stimuli even though the stimuli had reduced stress cues. Importantly, however, this comparison suggests that participants use lexical stress information in word learning only when they are forced to do so. Stress patterns may thus be used in the representation and recognition of new words, but only when the situation highlights that prosodic information is necessary to distinguish between segmentally identical stimuli.

In Italian, as in many other languages, the main difference between words is at the segmental level: Words mostly differ in their consonants and vowels [6; 12; 15]. If listeners are aware of this, then they may initially focus on phonetic cues

to segments during word learning. Although minimal stress pairs exist in Italian (e.g., *ANcOra* vs. *anCOra*), they are few in number. For this reason, if listeners are not forced to consider stress information, they have little reason to take suprasegmental cues into account when they hear a new word for the first time. Thus, even if all Italian listeners can easily distinguish between two well-known words that differ only in stress pattern (*ANcOra* vs. *anCOra*), our results suggest that they will have a problem when doing the same with newly-learned words. That is, if they are not explicitly made aware of the fact that two previously unknown words differ only in stress, then they are likely to represent those new items as not being phonologically different. To interpret two new stimuli as two really different words, listeners need the new words to be different at the segmental level, that is, in their consonants or vowels. This is because stress differences appear not to be automatically encoded in new lexical representations.

## 5. Acknowledgments

We thank Giusy Turco for recording the stimuli.

## 6. References

- [1] Dahan, D., Magnuson, J.S., Tanenhaus, M.K., and Hogan, E. M., "Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition", *Lang. Cognitive Proc.*, 16: 507-534, 2001.
- [2] Norris, D., and McQueen, J.M., "Shortlist B: A Bayesian model of continuous speech recognition", *Psychol. Review*, 115: 357-395, 2008.
- [3] Reinisch, E., Jesse, A., and McQueen, J.M., "Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately", *Quart. J. Exp. Psychol.*, 73: 772-783, 2010.
- [4] Sulpizio, S., and McQueen, J.M., "Italians use abstract knowledge about lexical stress during spoken-word recognition", submitted.
- [5] Magnuson, J.S., Tanenhaus, M.K., Aslin, R.N., and Dahan, D., "The time course of spoken word learning and recognition: Studies with artificial lexicon", *J. Exp. Psychol.: Gen.*, 132: 202-227, 2003.
- [6] Toro, J.M., Nespore, M., Mehler, J., and Bonatti, L.L., "Finding words and rules in a speech stream", *Psychol. Science*, 19: 137-144, 2008.
- [7] Shatzman, K.B., and McQueen, J.M., "Prosodic knowledge affects the recognition of newly acquired words", *Psychol. Science*, 17: 372-377, 2006.
- [8] Creel, S.C., Tanenhaus, M.K., and Aslin, R.N., "Consequences of lexical stress on learning an artificial lexicon", *J. Exp. Psychol.: Learn., Mem. and Cogn.*, 32: 15-32, 2006.
- [9] Thornton, A.M., Iacobini, C., and Burani, C., "*BDVDB Una base di dati sul vocabolario di base della lingua italiana*" [BDVDB: A database for the Italian basic dictionary]. Roma: Bulzoni, 1997.
- [10] Krämer, M., "The phonology of Italian", New York: Oxford University Press, 2009.
- [11] Boersma, P., "PRAAT, a system for doing phonetics by computer", *Glott International*, 5:341-345, 2001.
- [12] Nazzi, T., and New, B., "Beyond stop consonants: Consonantal specificity in early lexical acquisition", *Cog. Develop.*, 22: 271-279, 2007.
- [13] Curtin, S., "Twelve-month-olds learn novel word-object pairings differing only in stress pattern", *J. Child Lang.*, 36: 1157-1165, 2009.
- [14] Curtin, S., "Young infants encode lexical stress in newly encountered words", *J Exp. Child Psychol.*, 105: 376-385, 2010.
- [15] Nazzi, T., and New, B., "Use of phonetic specificity during the acquisition of new words: differences between consonants and vowels", *Cognition*, 98: 13-30, 2005.