Cues to lexical stress assignment in reading Italian: A megastudy with polysyllabic nonwords

Giacomo Spinelli[1], Sonia Trettenero[2], Stephen J. Lupker[3], Lucia Colombo[2]

[1]University of Milan-Bicocca, Italy

[2]University of Padua, Italy

[3]University of Western Ontario, Canada

Running Head: A megastudy with polysyllabic nonwords in Italian

Corresponding Author:

Giacomo Spinelli

Dipartimento di Psicologia, Università degli Studi di Milano-Bicocca

Piazza dell'Ateneo Nuovo 1, 20126 Milano (MI), Italy

E-mail: giacomospinelli@hotmail.it

Word count: 20528

Abstract

When reading polysyllabic words, assignment of lexical stress is a challenge for readers, especially in languages, such as English or Italian, in which stress position is not strictly determined even though words as well as nonwords typically contain several sublexical cues to stress that readers might use. Here, we attempted to identify such cues using a corpus analysis and to examine their impact on human performance in a megastudy in which participants (*N* = 45) assigned stress to nonwords (*N* = 800), stimuli particularly revealing of stress cue use because they have no predefined stress pattern. Hierarchical regression results confirmed an impact of sublexical cues examined in former studies and revealed a role for cues not previously examined, including similarity to real words. These results are informative for computational models of reading as they indicate that readers assign stress to nonwords based on not only sublexical but also lexical information.

Keywords: lexical stress; stress assignment; Italian; reading; megastudy

Cues to lexical stress assignment in reading Italian: A megastudy with polysyllabic nonwords

Since the 1970s many theoretical and computational models have been proposed in the literature on word recognition in an effort to explain how letters, graphemes, and words are processed (e.g., Coltheart et al., 1977; McClelland & Rumelhart, 1981; Seidenberg & McClelland, 1989). However, the way in which lexical stress (i.e., the emphasis placed on a syllable within a polysyllabic word that makes that stressed syllable more acoustically prominent) is assigned and how the process of stress assignment interacts with other word recognition processes have been relatively less explored issues. As a result, it is not clear to what extent stress assignment is a process that is driven by various aspects of letter processing or by lexical retrieval. One of the main reasons why such is the case is that most early word recognition models were developed for monosyllabic English words (e.g., Coltheart et al., 2001; Perry et al., 2007; Plaut et al., 1996).

Some exceptions to this trend, however, do exist for English, specifically the Connectionist Dual-Processing (CDP++) model by Perry et al. (2010), Ševa et al.'s (2009) connectionist model (SMA09), and, even earlier, Rastle and Coltheart's (2000) rule-based model (RC00). For languages other than English, computational studies and modelling have been rarer still. For example, for Italian, the language investigated here, only two models have been developed, one by Pagliuca and Monaghan (2010) and the other by Perry et al. (2014).

Computational models of this sort have typically been proposed in an attempt to simulate relevant human experimental data reported in the literature, particularly data from reading aloud studies in which polysyllabic real words and/or nonwords are assigned a specific stress pattern (e.g., first-syllable stress) when being pronounced. These empirical studies have helped shed some light on the nature of the processes involved in stress assignment. For example, Colombo's (1992) seminal experiments made the case for Italian that stress is assigned by different processes depending on the nature of the stimulus to be read. Italian is a transparent language at the segmental level (i.e., phonemes can be easily derived from print in most situations) but not at the suprasegmental level (i.e., stress patterns cannot be easily derived from print in most situations). For known words, particularly high-frequency words, stress information would be most quickly retrieved by looking up their representations in the mental lexicon. For low-frequency words and nonwords, however, another process would be involved. That process would be a sublexical one based on distributional information learned by readers during their experience with the Italian language.

There are two different ways to conceive of what is learned (and, then, used) in this process. One is that readers would learn rules, i.e., deterministic print-to-stress mappings. For example, in several languages (including Italian), one stress pattern—the dominant stress pattern—is more frequent than others (e.g., in Italian, the penultimate, or second-from-last, pattern, as in "piSTOla" 'gun'; henceforth, the stressed syllable is capitalized in all of our examples). According to some production models (e.g., Levelt et al., 1999; Schiller et al., 2004), the rule would then be learned that a word/nonword in that language should always be assigned the dominant stress pattern unless some other lexical or sublexical process interferes with this default assignment. (Note 1)

More recently, however, the view has been favored that what is being learned and used in the process of stress assignment is *probabilistic* print-to-stress mappings. For example, for Italian, readers would be able to learn that penultimate stress is the most probable stress pattern. However, they would also be able to learn probabilistic associations between orthographic endings and their typical stress pattern, associations known as stress neighborhoods. For example, "-ino" is a dominant stress neighborhood (i.e., many words ending in "-ino" bear stress on the penultimate syllable, the dominant stress pattern) whereas "-ico" is an antepenultimate stress neighborhood (i.e., many words ending in "-ico" bear stress on the antepenultimate syllable, a nondominant stress pattern; see below for a more complete definition of stress neighborhood). All of these learned associations (and, potentially, others) would be used when assigning stress (for further information on the characteristics of Italian and potential cues to stress in that language, see Study 1 below).

Several empirical studies (both in Italian and in other languages) have focused on this type of *statistical learning* (i.e., learning of probabilistic print-to-stress mappings which cannot be described by rules or with reference to individual representations in memory; see, e.g., Arciuli & Cupples, 2006; Ktori et al., 2018; Treiman et al., 2020). The majority of these studies, particularly the Italian ones, have used the standard factorial approach, in which researchers typically select two specific variables, such as type of stress pattern (e.g., first-syllable vs. second-syllable stress) and type of stress neighborhood (e.g., the typical vs. atypical stress pattern for words with the same ending), and then collect a set of items that fit the (most typically, four) cells of the experimental design.

More recently, a few *megastudies*, a procedure involving a complementary approach to the factorial approach, have appeared in the literature. In megastudies, a large number of stimuli are selected, varying on many dimensions, allowing several potential sources of information to be simultaneously investigated without the limitations of factorial designs. In stress assignment research, megastudies are useful because, first, they can provide a comprehensive picture of not only the various processes involved in stress assignment but also a number of more general reading processes. Second, they can provide a large database of stimuli and responses. Finally, they can provide answers to theoretically important questions, answers which would be hard for factorial studies to provide. Jouravlev and Lupker (2015b) and Mousikou et al. (2017) were among the first to apply this alternative approach to research on the stress assignment process, although their research questions differed as did their procedures.

Jouravlev and Lupker (2015b) used the megastudy approach to examine the sensitivity of Russian readers to potential stress cues that are available in the Russian vocabulary. Accordingly, their procedure first included a corpus analysis of 13,945 disyllabic Russian words in order to test the reliability of potential stress cues such as the word’s grammatical category, frequency, and existing associations between different types of orthographic units and stress. In a subsequent non-factorial study (i.e., a megastudy) in which 500 words were presented to 34 Russian participants to read aloud, three orthographic sublexical components (i.e., the stress neighborhoods associated with the entire first syllable and the entire second syllable and that associated with the ending of the second syllable) were found to be used as cues to stress by Russian readers in a regression analysis. These findings, which were further confirmed in a factorial study with a more limited number of words, suggest that, to at least some extent, readers learn and use statistical regularities in their language (see also, e.g., Arciuli & Cupples, 2006; Treiman et al., 2020, 2023).

Mousikou et al. (2017) used the megastudy approach to contrast competing computational models with respect to their ability to simulate human performance (for a similar procedure using a factorial approach, see Ktori et al., 2018). To do so, Mousikou et al. first tested 41 adult English readers, whose task was to read aloud a set of 915 disyllabic nonwords. They used nonwords in order to allow them to investigate sublexical processing, assuming (within a dual-route reading framework) that the lexical pathway would play little, if any, role in reading nonwords. They found, through item-level regression analysis, that participants' performance was affected by several types of sublexical units. In particular, English readers tend to assign stress to the first syllable if the onset and rime units of the syllables making up the nonword are consistently associated with first-syllable stress within the English lexicon or if the first syllable of the nonword is orthographically heavier (i.e. has more letters in it) than the second syllable or if the nonword does not have orthographic neighbors with second-syllable stress. (Note that there were more nonwords that were assigned first-syllable stress than nonwords assigned second-syllable stress in their stimulus set.)

Mousikou et al. (2017) then used those results to explore how the three main computational models of polysyllabic English reading (i.e., the CDP++ model by Perry et al., 2010, the connectionist model by Ševa et al., 2009 [SMA09], and the rule-based model by Rastle and Coltheart, 2000 [RC00]) fared in simulating human performance. When the same sample of nonwords presented to the human participants was presented to those models, all were fairly good at simulating the human data. In particular, the CDP++ predicted 81% of the human responses, the SMA09 79%, and the RC00 73%. Because the performance of the CDP++ and the SMA09 was more similar to the human data than that of the RC00, Mousikou et al. concluded that the statistical-learning approach used to implement the CDP++ and SMA09 models (as opposed to the rule-based approach implemented by the RC00) might provide a better representation of human behaviour. (Nonetheless, the two superior models still had some limitations, namely, the ability to process only disyllabic units and their overgeneralization of the spelling-to-sound associations learned during the training phase.) Some of these cues and others were then further examined in factorial studies by Ktori et al. (2018) and Treiman et al. (2020), with their results providing further support for statistical-learning models of stress assignment, as opposed to rule-based models.

In the present research we attempted to expand on this work by applying a similar combined procedure (i.e., one involving the megastudy approach as well as a corpus analysis and a comparison with a computational model) to Italian, a language in which the findings/conclusions of the work reviewed thus far may not necessarily apply.

Many studies in Italian have investigated, through factorial experiments and corpus analyses, the characteristics of Italian stress assignment and several types of cues to stress that Italian readers do use (Burani & Arduino, 2004; Burani et al., 2014; Colombo, 1992; Colombo et al., 2014; Colombo & Sulpizio, 2015, 2021; Colombo & Zevin, 2009; Spinelli et al., 2016; Sulpizio et al., 2013; Sulpizio & Colombo, 2013) or might use (Ševa et al., 2009; Spinelli, Sulpizio, et al., 2017; Sulpizio et al., 2017; for a review, see Sulpizio et al., 2015). Some of those studies have also been used as a benchmark for the two computational models of Italian polysyllabic reading, Pagliuca and Monaghan's (2010) connectionist model and the Italian version of the CDP++ (Perry et al., 2014). Although both models involve stress processing, neither had complete success in simulating the patterns suggested by the human data in the relevant experiments.

For example, Pagliuca and Monaghan's (2010) model, a classic parallel distributed processing (PDP) model with the added ability to process stimuli up to three syllables in length, did not fully reproduce the pattern produced by humans in the one experiment relevant to stress for which a simulation was reported. Specifically, whereas the original experiment, Burani and Arduino's (2004) Experiment 1, produced an overall stress-neighborhood effect in the reading latencies, i.e., faster latencies for words having a stress pattern consistent with their neighborhood (e.g., "COmico" 'comical', which bears the stress pattern associated with "-ico", i.e., the antepenultimate one) than for words having a stress pattern inconsistent with their neighborhood (e.g., "neMIco" 'enemy', which does not bear the stress pattern associated with "-ico"), Pagliuca and Monaghan's (2010) model produced a similar pattern only in the error rates, and only with words bearing a nondominant stress pattern.

CDP++ Italian (Perry et al., 2014), a dual-route model of reading involving computation of stress information via both representations at the lexical level and/or learned print-to-stress associations at the sublexical level, fared better overall as it managed to simulate the patterns documented in a number of experiments relevant to stress. For example, it did reproduce the pattern of results for Burani and Arduino's (2004) Experiment 1, the experiment Pagliuca and Monaghan's (2010) model had difficulty with. However, it failed to reproduce the results of an additional experiment of Burani and Arduino's (2004), their Experiment 2. That experiment focused on the impact of the *number* of words in the stress neighborhood consistent with the word's stress pattern, the so-called "stress friends" (e.g., if the word is "COmico", a stress friend would be "CInico" 'cynical', whereas 'neMIco' would be a "stress enemy"). Further, CDP++ Italian had difficulty capturing the patterns of a few other experiments (for a more in-depth discussion, see Study 3 below).

The present megastudy, informed by a corpus analysis and providing one of the most comprehensive examinations of stress assignment in Italian, can serve as a new, fundamental benchmark for computational models of polysyllabic reading in Italian as well as provide some insight into past and new challenges for those models. For example, it is possible that some of the difficulty CDP++ Italian experienced with simulating the impact of the number of stress friends for Burani and Arduino's (2004) Experiment 2 might be due to the specific stimuli that Burani and Arduino used (do note that different stimuli have been known to produce different patterns in relevant research, and in fact, Burani and Arduino's Experiment 1 produced a different pattern of results than Colombo's (1992) Experiment 2 despite the fact that the two experiments involved similar manipulations but, critically, different stimuli). Examining that pattern again in the context of a megastudy, i.e., a study with a larger number of stimuli and in which the distribution of the relevant stress cues is a close match to that of the language, would help clarify whether that failure of CDP++ Italian's was an idiosyncratic one or one indicative of a fundamental model limitation. The present megastudy would also help with theoretical issues that have yet to be addressed, such as the overlap between morphological and orthographic representations in stress-neighborhood effects, an issue that would be hard to examine with a factorial study and, in fact, has not been examined thus far (for further details, see Study 1 below).

The present manuscript is organized as follows. We first report a corpus analysis guided by the literature on stress assignment (and the processing of stress patterns in general) in Italian in order to determine a set of cues that seem to be related to stress patterns in that language (Study 1). Our assumption was that, based on statistical-learning principles, the cues we could identify were potentially usable by Italian readers as probabilistic associations (as opposed to rules). Once the basis for those cues was established through the corpus analysis, we move on to report a nonword reading aloud megastudy (Study 2). Using a nonfactorial design and a large number of nonword stimuli, this megastudy aimed to examine, first, the impact of different sublexical cues to stress available in the vocabulary; and second, the impact of *lexical* information, information that is not the result of learning statistical regularities (or rules). We used nonwords because we assumed, in line with Mousikou et al. (2017), that the impact of cues on stress assignment would emerge more clearly with stimuli which do not have a stress pattern of their own but must be assigned one when they are read aloud. In addition, we created nonwords with a relatively large length range (3–5 syllables), thus overcoming one limitation of former studies (particularly the English ones), in which the length range examined was typically narrow (e.g., only disyllabic stimuli were examined). Finally, we contrast the results of the megastudy with those of a simulation conducted with CDP++ Italian (Perry et al., 2014) in order to gain some insight into the merits and potential challenges that computational models of Italian reading might experience when it comes to modelling stress assignment (Study 3).

Study 1: Corpus analysis

We begin by discussing the general characteristics of Italian with respect to stress. We then review the sublexical cues to stress highlighted by the relevant literature and, finally, we describe how we examined the impact of those cues in the present corpus analysis.

Characteristics of Italian

Italian is a free-stress language and, according to the source of our corpus, the Phonitalia database (Goslin et al., 2014; see below for a description of the corpus), the vast majority (i.e., over 98%) of Italian words are polysyllabic. Italian polysyllabic words can exhibit four kinds of stress patterns, which are classified starting from the final syllable backward: the final (last) syllable stress pattern (e.g., “coliBRÌ” ‘hummingbird’), the penultimate (second-from-last) syllable stress pattern (e.g., “piSTOla” ‘gun’), the antepenultimate (third-from-last) syllable stress pattern (e.g., “TAvolo” ‘table’), and the pre-antepenultimate (fourth-from-last) syllable stress pattern (e.g., “Abitano” ‘they live’). The frequencies of these four stress patterns in the language differ considerably. Disyllabic words are almost always stressed on the penultimate syllable. With more than two syllables, the situation is more complex, but the dominant pattern, as noted, is still penultimate syllable stress. Indeed, about 77% of Italian words, including all syllable lengths except for monosyllabic words, are stressed on the penultimate syllable and only 18% are stressed on the antepenultimate syllable with the remaining 5% involving words with other stress patterns (Spinelli, Sulpizio et al., 2017; see also Thornton et al., 1997).

As also noted, although the Italian language is highly transparent at the segmental level, stress information is neither entirely predictable based on any explicitly taught rule nor is it orthographically marked in most cases. Still, there are cases in which the stress position is virtually certain. First, when stress is placed on the word’s final syllable, it is orthographically marked by a diacritic on the vowel (e.g., “coliBRÌ” ‘hummingbird’). Second, when the word’s penultimate syllable is heavy (i.e., when it ends with a consonant, e.g., “concerto” ‘concert’), it virtually always bears stress, with some rare exceptions such as “MANdorla” ‘almond’ and

“POlizza” ‘policy’. Another exception is when an antepenultimate stress infinitive is combined with a clitic, e.g., “PRENdere” ‘to take’ + “te” ‘you’ = “PRENderti” ‘to take you’, in which case the clitic does not modify the infinitive’s original stress pattern.

Cues to stress in Italian

It has been shown that skilled Italian readers rely on both lexical and sublexical sources of information to decide the position of stress when reading aloud isolated polysyllabic words, with sublexical sources of information also being used when reading aloud polysyllabic nonwords. Which cues to stress play a role in this process and the relative importance of each cue are the issues at the core of the present investigation.

To date, researchers addressing the issue of stress assignment in Italian reading aloud (Burani & Arduino, 2004; Burani et al., 2014; Colombo, 1992; Colombo et al., 2014; Colombo & Sulpizio, 2015, 2021; Colombo & Zevin, 2009; Sulpizio et al., 2013; Sulpizio & Colombo, 2013) have focused on the three main cues to stress identified by Colombo (1992) and briefly discussed above. One of these, word knowledge, is a lexical one and has been assumed to be quickly available only when processing high-frequency words. The other two cues, the distribution of stress patterns in Italian (stress dominance) and the distribution of the stress patterns of words sharing the same ending as the word or nonword (stress neighborhood; more details below), are sublexical and would be involved in reading low-frequency words and nonwords. However, other sources of information have also been investigated, such as the morphosyntactic properties of word and nonword stimuli (Spinelli et al., 2016). Further, distributional analyses suggest a potential role for yet other sources of information, such as information conveyed by word beginnings or by the individual syllables making up a word (Spinelli, Sulpizio et al., 2017; Sulpizio et al., 2017).

Stress-neighborhood information

For any given word or nonword stimulus, stress neighborhood has been defined (Colombo, 1992) as the set of words having the same orthographic ending as that stimulus (i.e., the graphemes going from the vowel of the penultimate syllable to the end of the stimulus, which most typically—i.e., with VCV endings—correspond to the last three letters). Stress neigbhorhood is known to influence stress assignment. For example, when the orthographic ending of a word/nonword stimulus belongs to a stress neighborhood in which there is a majority of dominant (i.e., penultimate) stress words (e.g., "-ino" as in "bamBIno" 'child'), the probability of assigning the dominant stress pattern to that stimulus is high. Two aspects of stress neighborhood have been found to be influential in stress assignment in Italian: numerosity or count (the *number* of words in the neighborhood having the same stress pattern) and consistency or proportion (the *proportion* of words in the neighborhood having the same stress pattern: Burani & Arduino, 2004; Burani et al., 2014; Colombo, 1992; Colombo et al., 2014; Colombo & Sulpizio, 2015; Colombo & Zevin, 2009; Sulpizio et al., 2013; Sulpizio & Colombo, 2013).

Note that stress-neighborhood count and proportion are potentially dissociable in Italian because, although both large and small neighborhoods can have high proportions of, for example, dominant stress words, the dominant-stress word count in the former would be much higher than that in the latter. More importantly for present purposes, the contributions of stress-neighborhood count and proportion in Italian have been shown to be empirically independent of one another (Burani & Arduino, 2004; Sulpizio et al., 2013). It has also been shown that type measures of counts and proportions (i.e., not frequency-weighted measures) outperform token measures (i.e., frequency-weighted measures; Sulpizio et al., 2013). Because of this finding as well as similar results from other languages (e.g., Jouravlev & Lupker, 2015b), all of the count and proportion measures used in the present research are type measures.

In the present analysis, we examined the impact of stress neighborhood using, for each word in the corpus, the number and the proportion of dominant stress neighbors for that word (henceforth, *dominant-stress-neighborhood count* and *dominant-stress-neighborhood proportion*, respectively) as predictors of the word's stress pattern. The expectation was for higher counts and proportions to be associated with a higher likelihood for the word to bear dominant stress.

Syllabic information

Another set of predictors was related to syllabic information, derived from STRESYL (Sulpizio et al., 2017), a database documenting different degrees of association between specific syllables and stress. Using this database, we determined, for each syllable of the stimuli we used, the number and proportion of cases in which that syllable bears stress within the Italian lexicon *when appearing in the same position* in a word. For example, the nonword “bosami” is formed by the syllables “bo”, “sa”, and “mi”. If we consider the penultimate syllable “sa”, we find that it bears stress in penultimate position in a relatively large number (932) and high percentage of Italian words (i.e., 90%, calculated in relation to all the words with the syllable “sa” in penultimate position). And if we consider the antepenultimate syllable “bo”, we find that it bears stress in antepenultimate position in a relatively small number (43) and low percentage (14%) of Italian words. Overall, then, this particular combination of syllables (i.e., “bo” in the antepenultimate position and “sa” in the penultimate position) would seem to favor a penultimate stress pattern, assuming that Italian readers use such information. Although Sulpizio et al. (2017) documented varying degrees of association between syllables and stress, they did not conduct regression analyses to examine their impact while controlling for the impact of other relevant predictors such as stress neighborhood.

In contrast, we did so in the present corpus analysis using four variables as predictors of the word’s stress pattern (for a similar analysis, see Jouravlev & Lupker, 2015b). The first two are the *penultimate-syllable stress count* and the *penultimate-syllable stress proportion*, that is the number and proportion of words in which the word’s penultimate syllable bears stress when appearing in the penultimate position, respectively. The second two are the *antepenultimate-syllable stress count* and the *antepenultimate-syllable stress proportion*, that is the number and proportion of words in which the word’s antepenultimate syllable bears stress when appearing in the antepenultimate position, respectively. (We did not consider syllables in other positions because the preantepenultimate position rarely receives stress and the final position never does unless it is marked with a diacritic, a type of final syllable we have excluded from the present research.)

We expected that, with high penultimate-syllable stress count and/or proportion (i.e., when the word's penultimate syllable occurs in a relatively large number and/or high proportion of words stressed in the penultimate position), there should be a higher likelihood for the word to bear the dominant (penultimate) stress pattern. By the same token, with high antepenultimate-syllable stress count and/or proportion (i.e., when the word's antepenultimate syllable occurs in a relatively large number and/or high percentage of words stressed in antepenultimate position), there should be a lower likelihood for the word to bear the dominant (penultimate) stress pattern. Note that the two sets of variables are not the inverse of each other because the set of words that share the word's penultimate syllable (e.g., a "sa" in penultimate position) will never completely overlap with the set of words that share the word's antepenultimate syllable (e.g., a "bo" in antepenultimate position). Therefore, these two factors should be relatively independent of one another. Overall, we expected that once the effect of the stress neighborhood is partialled out, considering which specific syllable occupies the penultimate or antepenultimate position might explain additional variance in the distribution of stress patterns in Italian.

An additional predictor that we included was length in syllables, i.e., the number of the syllables in the nonword. In Italian, the dominant stress pattern is not as common in longer words as it is in shorter words: For example, while about 81% of three-syllable words bear the dominant stress, that percentage drops slightly to 79% for four-syllable words and to 71% for five-syllable words (Spinelli, Sulpizio et al., 2017). Again, these descriptive data have not yet been corroborated with regression analyses demonstrating an association between length in syllables and stress patterns while controlling for the impact of other relevant predictors. The present corpus analysis can provide such a demonstration. (Note 2)

A syllabic predictor we chose not to include is the penultimate syllable weight. As noted, a heavy penultimate syllable (e.g., "concerto" 'concert'), virtually always receives stress in Italian. Indeed, words with a heavy penultimate syllable in our corpus had penultimate stress in 95.83% of cases, with almost all of the remaining words belonging to the antepenultimate stress infinitive + clitic case noted above (e.g., "PRENderti"). Therefore, we did not feel a need to demonstrate this fact in our regression analysis and we focused on words with a light penultimate syllable instead.

Morphological information

A large body of linguistic literature has investigated the role of morphemes in word recognition. In particular, psycholinguistic research has focused on the issue of how morphemes are represented and whether morphemically complex words are decomposed in reading and word recognition (Andrews, 1986; Burani & Caramazza, 1987; Burani et al., 1984; Marslen-Wilson et al., 1994; Napps, 1989; Taft, 1979; Taft & Forster, 1975; see also Lupker & Spinelli, 2023), with one influential view being that morphemes have an independent linguistic status compared to semantics and phonology. A contrasting view proposed in the framework of connectionist models was that the morphological structure is an emergent property reflecting “the systematic though probabilistic relationships among phonological, orthographic and semantic codes” (Gonnerman et al., 2007). However, if morphemes, particularly affixes, have an independent status from orthography and phonology and have stable associations with stress patterns, they could be also used by readers as cues to stress, in addition to stress-neighborhood and syllabic information.

Evidence from English does suggest that affixes are important cues for stress assignment (Arciuli & Cupples, 2006, 2007; Ktori et al., 2016; Mousikou et al., 2017; Rastle & Coltheart, 2000; Treiman et al., 2020), with orthographic sequences corresponding to prefixes, in particular, repelling stress. (Note 3) The importance of affixes has been especially highlighted in Rastle and Coltheart’s (2000) model of English dysyllabic word reading. That model was formulated within the dual-route framework and consisted of a lexical look-up procedure as well as a sublexical procedure based on spelling-to-sound rules. In order to predict the assignment of stress to English disyllables, Rastle and Coltheart proposed an algorithm consisting of a set of grapheme-to-stress rules based on the identification of prefixes and suffixes. The algorithm was quite successful at predicting English stress, in particular with prefixed words. However, it was not as successful at capturing the impact of other potential cues to stress, such as vowel length and orthographic weight (Ktori et al, 2016; Mousikou et al, 2017).

In contrast with English, in Italian, the role of affixes in stress assignment has not been investigated independently from orthography to this point. However, affixes are widespread in Italian (e.g., 85% of the words in the corpus of the present study have a suffix-like ending; see also Talamo et al., 2016) and have stable associations with stress patterns, suffixes in particular. For example, words with the suffix "-ato", a suffix of the past participle form of verbs with an infinitive ending in "-are" (i.e., first conjugation verbs), *always* bear the dominant (penultimate) stress pattern, while words with the suffix "-ico", a suffix used to form adjectives from nouns, *always* bear the nondominant, antepenultimate stress pattern. This fact may make suffixes potential stress cues distinct from stress-neighborhood cues (i.e., cues that do not involve such stable associations between spelling and stress patterns as suffixes do). Inevitably, however, these cues do correlate with stress-neighborhood cues (e.g., penultimate-stress and antepenultimate-stress suffixes are associated with stress neighborhoods in which penultimate and antepenultimate stress patterns prevail, respectively). By including both sets of predictors (the suffix ones and the stress-neighborhood ones), the present corpus analysis allowed an examination of their independent associations with stress patterns.

This examination is particularly important because suffix cues have often been confounded with stress-neighborhood cues in the studies that investigated Italian stress assignment. If suffixes turn out to be significant predictors on top of stress-neighborhood effects, their effect might be due to the application of either rules as has been proposed in Rastle and Coltheart's (2000) model or the learning of their especially strong associations with stress patterns. Some insight into the relevant processes might then be gained by comparisons with computational models that do not include specific rules in their system, such as the two models of Italian polysyllabic reading (Pagliuca & Monaghan, 2010; Perry et al., 2014). (Another possibility might be assuming that semantics plays a mediating role, however, this idea could not be verified using extant computational models because they have no access to semantics.)

In contrast, prefix cues were not included in our analysis because, although prefixes tend to repel stress in Italian as they do in other languages such as English, prefixes have no specific stress pattern associated with them. Therefore, in polysyllabic stimuli, the type of stimuli that we focused on, the presence of a prefix alone provides little information as to which syllable should be stressed, particularly when the stimulus has more than three syllables. Consistent with this idea, the frequency of dominant stress was similar for words with and without a prefix in our corpus (78.80% and 78.03%, respectively).

Having reviewed the information potentially associated with stress in Italian, next, we report an analysis of a corpus of polysyllabic Italian words that we conducted in order to determine the presence of such associations in Italian.

Data availability

For all analyses reported here, the raw data, R files used for the analyses, and study materials are available at https://osf.io/9q54v/. This research was not preregistered.

Method

We extracted our corpus from PhonItalia version 1.1 (Goslin et al., 2014), a lexical database providing orthographic and phonological information for 120,000 word forms with frequency counts and part-of-speech tagging derived from CoLFIS (Bertinetto et al., 2005). Frequencies ranged from 1 to 119,430 raw occurrences (excluding one word with a frequency of zero), that is from 0.3 to 36,036.7 occurrences per million. The version 1.1 of the database differs from the original version in that it includes corrections for 11,523 words for which the position of the stressed syllable was incorrect and for 79 words for which the phonological representation was incorrect. These corrections were suggested by the first author of the present manuscript (G. S.), who manually inspected the database, based on his knowledge of Italian and occasional consultation of the stress information available in the Sabatini-Coletti Italian vocabulary, accessed through https://dizionari.corriere.it/dizionario_italiano/.

The corpus was created as follows. First, PhonItalia lists word forms appearing within phrases (e.g., “night” in “Saturday night fever”) in different entries (i.e., rows) than the same word form (belonging to the same lemma and with the same part of speech) not appearing within phrases (e.g., “night”). Because this distinction was not relevant to the present research, we collapsed those entries and summed their associated frequencies.

Next, we selected words belonging to either the noun, verb, adjective, or adverb part of speech, which had 3 to 5 syllables and at least 5 letters, included no apostrophe (i.e., the symbol used in PhonItalia to indicate the stress diacritic), and ended with a consonant and a vowel (i.e., a Consonant-Vowel (CV) orthographic structure). The last three letters of the word (i.e., the ending) could thus have either a VCV or a CCV orthographic structure (the most common orthographic structures for endings in Italian, see Spinelli, Sulpizio et al., 2017). We used these constraints because they were the same as those that we applied to construct the nonwords for the megastudy (see Study 2 below) and we wanted the results of the present corpus analysis and those of the megastudy to be comparable. After applying these filters and removing words for which one of the measures of interest described below could not be calculated, the corpus included 71,492 words, with 0.01% bearing stress on the ultimate syllable, 78.46% on the penultimate syllable, 21.02% on the antepenultimate syllable, and 0.50% on the pre-antepenultimate syllable (percentages that are similar to those reported by Spinelli, Sulpizio et al., 2017, for the entire PhonItalia). Other characteristics of the words in the corpus are reported in Table 1.

Table 1

Characteristics of the words in the corpus

Characteristic	*M*	*SD*	Range
Frequency (occurrences per million)	4.12	20.70	.30–2,113.38
Length in letters	9.21	1.90	5–17
Length in syllables	3.89	0.74	3–5
N (orthographic neighborhood size)	2.74	3.37	0–31
Mean *N* frequency (occurrences per million)	2.88	7.97	0–298.42
Dominant-stress-neighborhood count	875.98	883.20	0–2,870
Dominant-stress-neighborhood proportion	.80	0.32	0–1
Penultimate syllable stress count	786.01	809.99	0–2,854
Penultimate syllable stress proportion	.78	0.23	0–1
Antepenultimate syllable stress count	99.52	109.38	0–461
Antepenultimate syllable stress proportion	.19	0.14	0–1

Note. Orthographic neighborhood size (*N*) and mean *N* frequency were calculated using the vwr package, version 0.3.0 (Keuleers, 2013), specifying the original PhonItalia database as the source for the orthographic neighbors.

With respect to the stress measures, we extracted the last 3 letters, operationally defined as the “ending”, and the counts and proportions associated with that ending in Q2Stress (Spinelli, Sulpizio et al., 2017). (Note 4) These were the *dominant-stress neighborhood count* (i.e., the number of dominant-stress words with that ending—on average, 875.98) and the *dominant-stress neighborhood proportion* (i.e., the proportion of dominant-stress words with that ending—on average, 78%). Likewise, for each word we extracted the penultimate and antepenultimate syllable and the counts and proportions associated with those syllables in STRESYL (Sulpizio et al., 2017). These were the *penultimate-syllable stress count* (i.e., the number of words in which the word’s penultimate syllable bears stress when appearing in the penultimate position—on average, 786.01), the *penultimate-syllable stress proportion* (i.e., the proportion of words in which the word’s penultimate syllable bears stress when appearing in the penultimate position—on average, 78%), the *antepenultimate-syllable stress count* (i.e., the number of words in which the word’s antepenultimate syllable bears stress when appearing in the antepenultimate position—on average, 99.52), and the *antepenultimate-syllable stress proportion* (i.e., the proportion of words in which the word’s antepenultimate syllable bears stress when appearing in the antepenultimate position—on average, 19%).

Finally, we determined whether the word included a potential prefix (*prefix presence*), dominant-stress suffix (*dominant-suffix presence*), or nondominant-stress suffix (*nondominant-suffix presence*). We did so using a look-up procedure, described in detail in the Supplementary Materials, which provides an approximation to the affixes the word may contain but is far from being error-free as it includes no verification that there are the necessary conditions for the word to potentially contain those affixes (a verification procedure which we did implement for the nonwords in our megastudy). The data that this procedure produced thus need to be taken with caution. That said, the procedure indicated a potential prefix in 39,889 words (55.80% of the corpus) and a potential suffix in 60,899 words (85.18%). Among the words with a potential suffix, that suffix was a suffix uniquely associated with the dominant stress pattern in the majority of the cases (49,341, with words including this type of suffix bearing dominant stress 91.02% of the time); it was a suffix uniquely associated with the antepenultimate stress pattern (i.e., a nondominant one) in 4,795 cases (with words including this type of suffix bearing dominant stress 4.84% of the time); and it could be associated with both dominant or nondominant stress patterns in 6,763 cases (with words including this type of suffix bearing dominant stress 33.05% of the time; note that this situation would occur for homographic suffixes such as “-ano”, which is associated with the dominant stress pattern when appearing in nouns and adjectives and with nondominant stress patterns when appearing in verbs).

Data Analysis

As noted, over 78% of the words in the corpus were stressed on the penultimate syllable. Those words were classified as bearing the dominant stress pattern, and the other words (mainly, antepenultimate-stress words) were categorized as bearing nondominant stress patterns. Stress pattern (dominant vs. nondominant) was then used as the dependent variable in a series of analyses using generalized linear modelling implemented with the glm function in R version 3.6.3 (R Core Team, 2020). Prior to the analysis, words with a heavy penultimate syllable were inspected. As noted, they had dominant stress in the vast majority (95.83%) of the cases. Those words (accounting for 18,884 observations) were discarded from the analysis, which was then based on a total of 52,608 observations.

Stress patterns were modelled specifying a binomial distribution and a logit link function (see, e.g., Sulpizio et al., 2013). Prior to running the models, R-default treatment contrasts were changed to sum-to-zero contrasts to help interpret lower-order effects in the presence of higher-order interactions (Levy, 2014). (While no interaction terms were included in the corpus analysis, they were in the subsequent analyses, see Studies 2 and 3 below.) Follow-up analyses were conducted using the emmeans package, version 1.4.6 (Lenth, 2018). Graphs were generated with the ggeffects package, version 1.0.1 (Lüdecke, 2018), the ggplot2 package, version 3.3.5 (Wickham, 2016), and the gridExtra package, version 2.3 (Auguie, 2017).

The inclusion of predictors in the model was determined using a hierarchical approach in which the predictors were entered in several steps, starting from the type of predictors that are most established for the dependent variable being examined and proceeding towards the least established ones (for similar approaches, see Boukadi et al., 2016; Cortese & Schock, 2013; Mailhot et al., 2020; Sánchez-Gutiérrez et al., 2018; Yap & Balota, 2009).

Predictors of stress patterns were entered in 4 steps. In Step 0, the *null model*, we entered the intercept only. In Step 1, we added the *stress-neighborhood predictors*: dominant-stress count and proportion in the stress-neighborhood. In Step 2, we added the *syllabic predictors*, that is, the predictors relevant to the syllabic structure of the nonword: length in number of syllables and penultimate and antepenultimate syllable stress count and proportion. Finally, in Step 3, we added the *morphological predictors*: dominant-stress suffix and nondominant-stress suffix, two predictors which were both dummy-coded as present vs. absent.

For each step, we dropped the predictors that did not reach statistical significance before proceeding to the next step. In all cases, dropping those non-significant effects resulted in a model that had (by definition) a simpler structure, but no worse goodness of fit (in terms of log likelihood) than the model with those effects, as demonstrated by a nonsignificant chi square test performed with the anova function in base R. For each step (excluding Step 0), the model without the non-significant predictors (i.e., the final model for that step) was also compared with the final model from the previous step with a chi square test to determine whether goodness of fit in the current step (i.e., the more complex model) had improved from the previous step (i.e., the less complex model). Estimates of R^2 (i.e., the proportion of total variance explained by the predictors) were also obtained for the final model in each step using the rsq function in the rsq package, version 2.5 (Zhang, 2022b).

Initially, the predictors were inspected for multicollinearity. To this end, we ran a model in which all predictors were entered simultaneously (e.g., Mailhot et al., 2020). On the basis of this model, we calculated the variance inflation factor for each of the predictors using the vif function in the car package, version 3.0-7 (Fox & Weisberg, 2019). This value was below 3 for all effects, posing little concern for multicollinearity. Furthermore, all the correlation coefficients for the (standardized) continuous variables used as predictors were less than .75 (Cohen et al., 2003). Therefore, no further adjustment for multicollinearity seemed to be required a priori (but see below). Tables reporting the correlations for the predictors used in this analysis as well as subsequent analyses are reported in the Supplementary Materials.

Results

The basic relationships between each individual predictor and the proportion of dominant-stress patterns are reported in Table 2 in the form of Pearson's correlations (for continuous predictors) and means (for categorical predictors). Further, for each predictor and for the interactions between the predictors that we examined, a plot of the relevant trend line (for continuous predictors) or marginal mean (for categorical predictors) along with its 95% confidence interval is displayed in Figure 1 for the stress-neighborhood predictors, in Figure 2 for the syllabic predictors, and in Figure 3 for the morphological predictors. These values were estimated with a model in which all predictors were entered simultaneously. The reason for the curvature in some of the lines is due to the lines being displayed in the response scale as opposed to the logit scale used for model estimation. For visualization purposes, the standardized continuous predictors were also back-transformed to the original scale. Note that the lines and means displayed in the figures can differ from the associations reported in Table 2 because they are estimated from models that control for the impact of the other predictors. Note, further, that Table 2 and Figures 1 to 3 are for descriptive purposes and include all predictors, irrespective of whether they would turn out to be significant.

Examination of the correlation values and graphs suggests a strong impact of stress neighborhood proportion and a reduced one for stress neighborhood count among the stress neighborhood predictors, with both impacts being in the expected direction (i.e., higher proportions and counts being associated with higher probability for the word to bear dominant stress). Among the syllabic predictors, the most noticeable impacts were those of the proportion predictors, i.e., penultimate-syllable and antepenultimate-syllable stress proportion, which were in line with the expectations for those predictors (i.e., increasing penultimate-syllable stress proportion being associated with a higher probability for the word to bear dominant stress and increasing antepenultimate-syllable stress proportion being associated with a lower probability for the word to bear dominant stress). Interestingly, the impacts of the corresponding count predictors were not only reduced but also were in the directions opposite to the expected ones, i.e., increasing penultimate-syllable stress count was associated with a *lower* probability for the word to bear dominant stress (Figure 2B) and antepenultimate-syllable stress count was associated with a *higher* probability for the word to bear dominant stress (Figure 2D). Note, however, that the basic relationships between the syllabic count measures and stress pattern reported in Table 2 had the expected signs (i.e., positive for the penultimate syllable, $r = .202$, and negative for the antepenultimate syllable, $r = -.096$).

A possible explanation for these seemingly unintuitive patterns is that, although the relevant proportion and count measures were not strongly correlated (for the penultimate syllable, $r = .370$, for the antepenultimate syllable, $r = .171$; see Table 2) and the variance inflation factors did not suggest multicollinearity issues, the presence of one measure (e.g., proportion) influenced the ability of the model to estimate the impact of the other measure (e.g., count). In order to examine this idea, we re-ran the model with all predictors entered simultaneously excluding, for the syllabic predictors, either the proportion ones or the count ones. The proportion predictors were virtually unimpacted by the absence of the corresponding count ones. For the count predictors, on the other hand, the absence of the corresponding proportion predictors did change the sign of the impact of penultimate-syllable stress count (from negative, as represented in Figure 2B, to positive) but did not change the sign of the impact of antepenultimate-syllable stress count (which remained positive as in Figure 2D). Overall, the impact of the syllabic count predictors does not appear to be particularly clear and, for this reason, we decided to exclude those predictors from the hierarchical regression analyses (i.e., we entered in the appropriate step only the syllabic proportion predictors, penultimate-syllable and antepenultimate-syllable stress proportion).

Finally, among the morphological predictors, both the presence of dominant-stress and nondominant-stress suffixes had pronounced impacts on the probability of penultimate stress assignment (see Table 2 and Figure 3).

Table 2

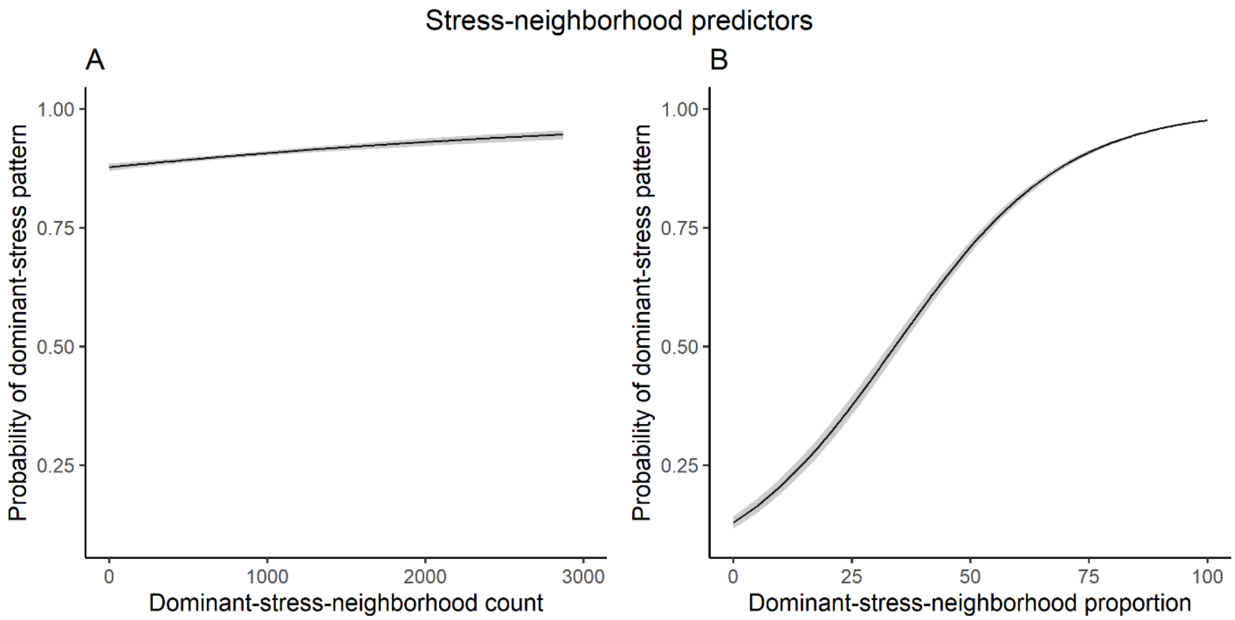
The basic relationships between predictors and proportion of dominant-stress patterns for the words used in the corpus analysis

Type of predictor	Predictor	Association with stress pattern
Stress-neighborhood	Dominant-stress-neighborhood count	r = .443
	Dominant-stress-neighborhood proportion	r = .819
Syllabic	Length in syllables	r = −.111
	Penultimate-syllable stress count	r = .202
	Penultimate-syllable stress proportion	r = .552
	Antepenultimate-syllable stress count	r = −.096
	Antepenultimate-syllable stress proportion	r = −.368
Morphological	Dominant-stress suffix presence	Absent: M = .41 Present: M = .80
	Nondominant-stress suffix presence	Absent: M = .85 Present: M = .19

Note. The critical value for significance (.05, two-tailed) for the correlations is r_c = .009.

Figure 1

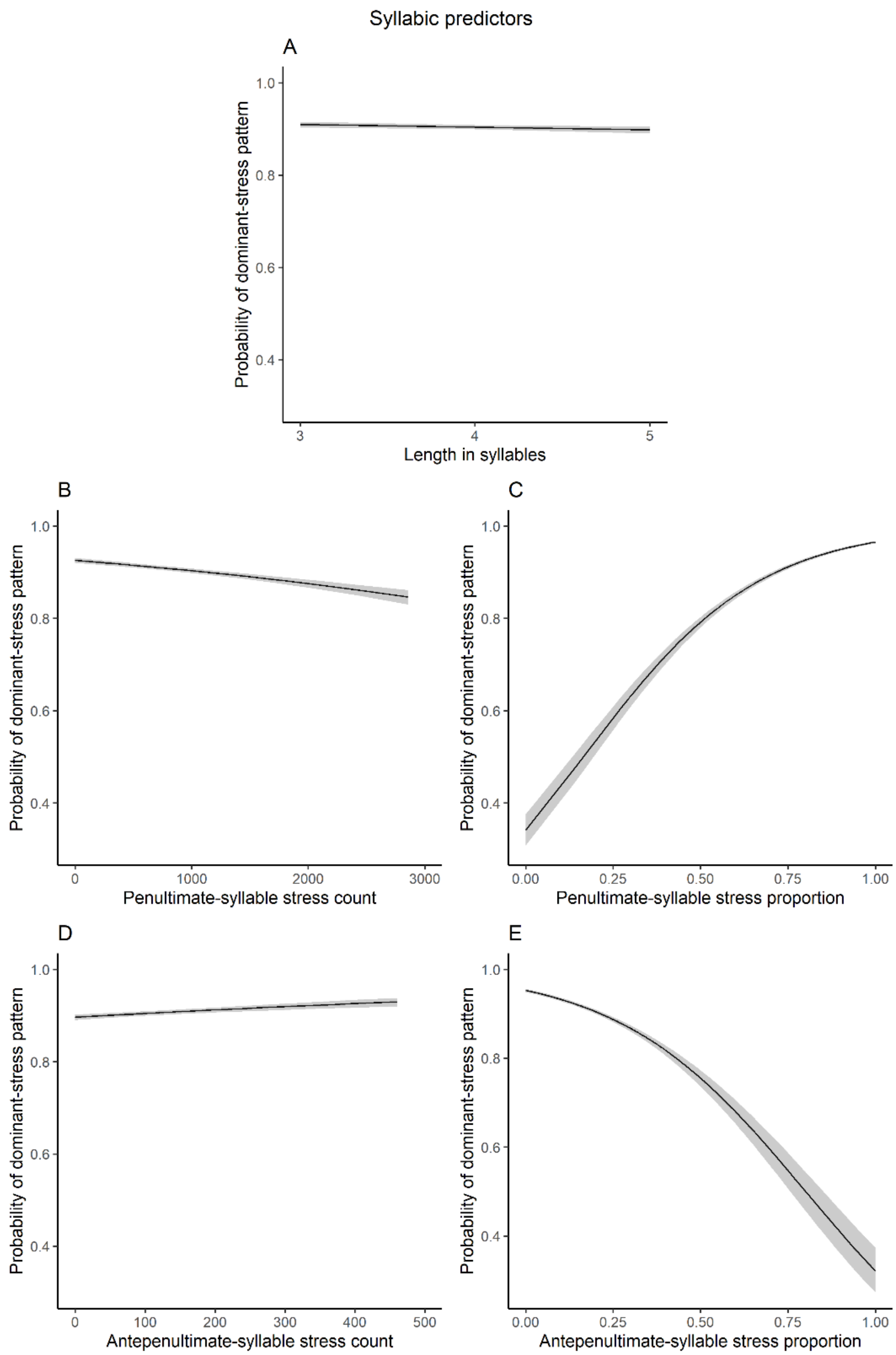
The impact of the stress-neighborhood predictors on the proportion of dominant-stress patterns for the words used in the corpus analysis



Note. The lines represent estimated intercepts and trends for the probability of dominant stress patterns as a function of dominant-stress-neighborhood count (A) and proportion (B).

Figure 2

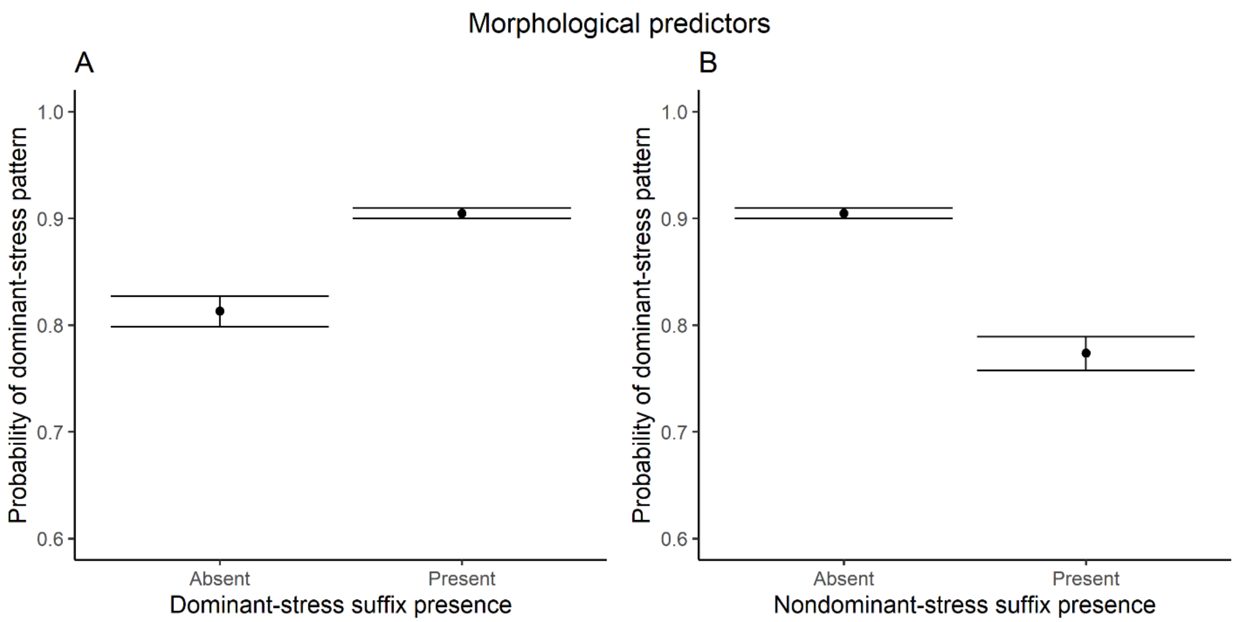
The impact of the syllabic predictors on the proportion of dominant-stress patterns for the words used in the corpus analysis



Note. The lines represent estimated intercepts and trends for the probability of dominant stress patterns as a function of length in syllables (A), penultimate-syllable stress count (B) and proportion (C), and antepenultimate-syllable stress count (D) and proportion.

Figure 3

The impact of the morphological predictors on the proportion of dominant-stress patterns for the words used in the corpus analysis



Note. The circles represent the estimated marginal means for the probability of dominant stress patterns as a function of the presence of a dominant-stress suffix (A) and nondominant-stress suffix (B).

The hierarchical regression results, reported in Table 3, are in general agreement with this assessment. In that table, we report, for each step, the predictors retained in the final model for that step, along with the parameters from the very final model in Step 3 (i.e., the best-fitting model). For each step, we also report the R^2 and the deviance value from the chi square test comparing the model in that step with the model from the previous step. Here, we describe the main patterns of results.

Table 3

Hierarchical regression results for the corpus analysis

		R^2	Model comparison		Best-fitting model parameters			
Step	Fixed effects		*Deviance*	*p*	β	*SE*	*z*	*p*
Step 1: Stress-neighborhood predictors		.6657	37,756	< .001				
	Dominant-stress-neighborhood count				0.217	0.038	5.74	< .001
	Dominant-stress-neighborhood proportion				1.94	0.024	81.94	< .001
Step 2: Syllabic predictors		.7079	3,498	< .001				
	Length in syllables				−-0.060	0.019	−3.07	.002
	Penultimate-syllable stress proportion				0.831	0.021	38.99	< .001
	Antepenultimate-syllable stress proportion				−-0.559	0.022	−25.35	< .001
Step 3: Morphological predictors		.7234	849	< .001				
	Dominant-stress suffix				−-0.388	0.023	−16.54	< .001
	Nondominant-stress suffix				0.538	0.023	23.27	< .001

Note. For each step, the model comparison is always between the model in that step and the model in the previous step. For the Step-1 model, the comparison is with the Step-0 model, i.e., the null model including the intercept only. Note that all continuous predictors were standardized prior to model evaluation.

First, for the *stress-neighborhood predictors* examined in Step 1, we confirmed a role for dominant-stress-neighborhood count and proportion: Words with an ending associated with a higher number and proportion of dominant-stress words in the lexicon were more likely to have dominant stress (see Figure 1), with the impact of the proportion of stress neighbors (Figure 1B) being stronger than that of the count (Figure 1A), as can be seen also by a comparison of the *r* and *z* values (Tables 2 and 3). It is also worth noting that when the stress neighborhood proportion is around 50% and thus essentially neutral, the estimate of the probability of dominant stress patterns is around 75% (Figure 1B), reflecting the fact that, even when stress neighborhood information involves no association with either stress pattern, the dominant stress pattern is still the most likely one.

Second, for the *syllabic predictors* examined in Step 2, we found a positive effect of penultimate-syllable stress proportion and a negative effect of antepenultimate-syllable stress proportion. These effects indicated that, as expected, words were more likely to have the dominant stress pattern when their penultimate syllable is one that attracts stress (i.e., it is more often stressed than unstressed; see Figure 2C) and less likely to have the dominant stress pattern when their antepenultimate syllable is one that attracts stress (see Figure 2E). Length in syllables also had an effect in the negative direction, with longer words being associated with a relatively lower probability for the word to bear the dominant stress pattern. This tendency is in line with the observation that dominant stress is slightly less frequent in longer words in Italian (Spinelli, Sulpizio et al., 2017). Note, however, that the effect is quite small, with the corresponding trend line in Figure 2A looking almost flat (the reason for the effect being significant is likely due to the large number of observations the model was based on).

Finally, for the *morphological predictors* examined in Step 3, we found effects of both dominant-stress and nondominant-stress suffixes. These effects indicated that, as expected, words with a dominant-stress suffix, compared to words without it, were more likely to have dominant stress (Figure 3A), and words with a nondominant-stress suffix, compared to words without it, were less likely to have dominant stress (Figure 3B). Note that because of the sum-to-zero contrasts we used and R's default alphabetical ordering of the levels of the suffix predictors (i.e., “n”, meaning the suffix is absent, followed by “y”, meaning it is present), it is normal for the sign of the dominant-stress suffix effect that we described to be negative and for the sign of the nondominant-stress suffix effect that we described to be positive (see Table 3; the same is true for the analyses reported below).

Discussion

As has been argued elsewhere (e.g., Arciuli & Cupples, 2006; Jouravlev & Lupker, 2015b), stress assignment may be driven by learned associations between sublexical units and stress patterns. The present corpus analysis allowed a determination of what those associations might be for Italian. We confirmed, through logistic regression, a role for several potential stress predictors previously highlighted in descriptive analyses (Spinelli, Sulpizio, et al., 2017; Sulpizio et al., 2017), particularly stress-neighborhood predictors (expressed both as counts and especially as proportions) and syllabic predictors (including length in syllables and proportion measures of associations between penultimate and antepenultimate syllables and stress). Further, both of the morphological predictors we used (i.e., dominant-stress and nondominant-stress suffix presence) were shown to have an impact above and beyond that of the other predictors, particularly that of predictors that are inevitably correlated with them such as the stress-neighborhood ones. That is, the presence of a dominant-stress suffix further increased, and the presence of a nondominant-stress suffix decreased, the likelihood of the word to bear dominant stress beyond the increase associated with the distributional information provided by the stress neighborhood (distributional information that, for words containing a dominant-stress suffix, typically favored dominant stress already). Indeed, inspection of Table 2 shows that the probability of dominant stress in a word is 80% when it includes a dominant-stress suffix, and 41% when it does not include it. An even stronger difference emerged for words including a nondominant-stress suffix: 19% when such a suffix is present and 85% when it is absent. In sum, stress neighborhood, a few syllabic predictors, and suffixes are correlated with stress, and may thus act as cues for humans to use in order to assign stress.

Study 2: Megastudy

The corpus analysis highlighted several potential associations between sublexical units and stress patterns Italian readers might learn. The questions we now move on to address are, first, do Italian readers actually use those associations when assigning stress, and second, is there a role also for lexical information (i.e., information less easily explained by statistical learning) in the stress-assignment process? We sought to answer these questions with a reading aloud megastudy involving nonwords created from words sampled from the corpus previously used.

In our megastudy, as noted, we used nonwords based on the assumption that the impact of cues to stress would emerge more clearly with stimuli which do not have a stress pattern of their own. Importantly, albeit it may sound counterintuitive, this assumption is true not only for sublexical cues but for lexical ones as well. Indeed, although some influential models of word recognition and reading aloud (e.g., Coltheart et al., 2001; Perry et al., 2007; Zorzi et al., 1998) assume that translation of graphemes into sounds for nonwords mainly occurs within a sublexical route, avoiding direct influences from the lexicon, other studies (e.g., Eddington, 2000; Glushko, 1979; Guion et al., 2003; Protopapas et al., 2007; see also Spinelli et al., 2016) have claimed that nonwords can also be read by analogy with similarly spelled words. Moreover, there is an argument that the lexical route is involved in the stress-neighborhood effects reviewed above. Perry et al. (2014) simulated those effects for Italian with the Italian version of the CDP++ model and explained them as being the result of an interaction between sublexical and lexical routes in the model.

Thus, if lexical information is indeed important, when nonwords are derived from existing Italian words, as in the present study, it can be hypothesized that the stress pattern of a nonword's source word would have some influence on stress assignment to that nonword. That is, when the source word has a dominant stress pattern, the probability of assigning a dominant stress pattern to the nonword can be expected to be higher, compared to when the source word has a nondominant stress pattern. To examine this idea, in the present megastudy we included among the predictors the stress pattern of the nonwords' source words and the degree of similarity of the nonwords to their source words, as any effect of the source word would likely be modulated by the degree of similarity existing between the nonword and its source word.

Specifically, nonwords were created by replacing a variable number of letters in a source word, allowing us to manipulate the degree to which the nonwords resemble their source word (e.g., one-letter replacement in a 10-letter word corresponds to a 10% change whereas a two-letter replacement in a 10-letter word represents a 20% change as does a one-letter replacement in a 5-letter word, etc.). We hypothesized that, the higher the similarity of the nonword to its source word, the stronger the influence of the stress pattern of the source word on the stress assignment to the nonword. Specifically, the (proportionally) fewer the replaced letters, the more activated the source word would be and, therefore, the stronger the effect of the source word's stress pattern. By a similar reasoning, we also included source word frequency information among the predictors, as we hypothesized that the more frequent the source word is, the more easily activated that word would be, and, therefore, the stronger its influence on performance.

The fact that the nonwords used in the present study were created from words is important also with respect to sublexical cues. Because the words used to create the nonwords (as will be explained below) were sampled at random from the corpus, we were able to create a situation in which the distribution of the sublexical cues contained by the nonwords more closely resembled the distribution of those cues within the original corpus and hence in the Italian lexicon (compared to factorial experiments in which the cues favoring infrequent stress patterns are typically overrepresented). Potential list context effects were thus minimized. Further, this design also allowed us to examine the relationship between stress assignment performance and continuous variation in some of the key stress predictors (e.g., stress neighborhood), predictors that have only been treated categorically in the past literature.

Our expectations for the sublexical cues were in line with the literature and/or the results of the present corpus analysis. Specifically, we expected, for the *stress-neighborhood predictors*, that nonwords with higher dominant-stress neighborhood counts and especially proportions would be more likely to receive dominant stress. For the *syllabic predictors*, we expected, first, that longer nonwords would be less likely to receive dominant stress; second, that nonwords with higher penultimate-syllable stress count and especially proportion (i.e., when the nonword's penultimate syllable occurs in a relatively large number and/or high proportion of words stressed in the penultimate position) would be more likely to receive dominant (penultimate) stress; and third, that nonwords with higher antepenultimate-syllable stress count and especially proportion (i.e., when the nonword's antepenultimate syllable occurs in a relatively large number and/or high percentage of words stressed in antepenultimate position) would be less likely to receive dominant (penultimate) stress. (Note that the corpus analysis did not produce clear evidence for the count measures for syllabic predictors, especially the antepenultimate-syllable stress count, to be effective stress cues. We examined those cues nonetheless because it is possible that count measures may have more weight than proportion measures on readers' stress assignment performance, see, e.g., Sulpizio et al., 2013.) Finally, for the *morphological predictors*, we expected that nonwords with a dominant-stress suffix and nonwords with a nondominant-stress suffix would be more and less likely, respectively, to receive dominant stress. Note that such a pattern of results with the morphological predictors, along with the expected patterns for the stress-neighborhood predictors, would produce the first demonstration that the effects of orthography and morphology in stress-neighborhood effects, normally confounded, are dissociable. Such a result would have also important implications for models of Italian polysyllabic reading, models which currently do not involve dedicated morphological or semantic systems (Pagliuca & Monaghan, 2010; Perry et al., 2014).

As with the corpus analysis, we expected little variability in stress assignment performance to nonwords with a heavy penultimate syllable. However, words with a heavy penultimate syllable are quite frequent in Italian (28%; see Spinelli, Sulpizio et al., 2017; Sulpizio et al., 2017). In order for our sample to approach the distribution of the two syllable types in Italian, 150 of the 800 the nonwords we created had a heavy penultimate syllable (i.e., 19%). These nonwords, however, were considered fillers in the analysis of stress responses, because we did not expect any variability in the assignment of stress to these stimuli.

Although our main focus was on the potential impact that the lexical and sublexical predictors reviewed above would have on nonword stress assignment (i.e., on *stress responses*), our study also allowed an examination of the variables influencing nonword reading aloud *latencies*. With respect to length, for example, it is known that longer stimuli take longer to name (Barca et al., 2002; Judica et al., 2002; Juphard et al. 2004; Zoccolotti et al., 1999) even though the length range examined in Italian studies has typically been quite narrow (from 5 to 9 letters long). Nonetheless, what happens when longer stimuli (words or nonwords) are presented is not clear. It is possible that the start of pronunciation will depend on the length of the letter sequence for stimuli in the 5–9 letter range, with latencies increasing linearly with every additional letter because readers may start articulation only once the stimulus has been fully processed. For longer stimuli, however, additional letters may have less of an impact because readers might tend to anticipate the articulation for most stimuli based on their initial letters. The length effect, in other words, may plateau for longer letter sequences, suggesting a quadratic component in that effect. Our stimuli, ranging from 5 to 13 letters, allowed an examination of these ideas (as well as other, typically unexplored ideas concerning the impact of stress cues on nonword reading latencies, such as whether cues that make the determination of the stress pattern easier also speed up latencies) through regression analyses similar to those conducted for stress responses. Because those analyses are not our main focus, their results are reported and discussed in the Supplementary Materials.

Method

Participants

Forty-five undergraduate students (average age = 24, range 21–28, *SD* = 2.8, 9 males) who were native speakers of Italian took part in the experiment as volunteers, a sample size that is comparable to that of Mousikou et al. (2017) (*N* = 41). They were asked to sign a written informed consent prior to their participation. All were right-handed, had normal or corrected-to-normal vision, and reported no reading impairments. This research was conducted in accordance with the Declaration of Helsinki.

Materials

Eight hundred nonwords were created following the phonotactic constraints of Italian. To create a nonword, first, a source word from the corpus previously described (see above) was selected at random based on the word's frequency (thus, more frequent words in the database were more likely to be selected than less frequent words). No word was selected twice. The mean length of the source words that were ultimately selected was 7.79 letters and 3.52 syllables. The mean frequency was 164.79 occurrences per million. The mean orthographic neighborhood size (*N*) was 4.50 and the mean *N* average frequency was 8.38 occurrences per million. Further, 556 of the source words (70%) had dominant stress, a percentage similar to the distribution of the dominant stress in the corpus and the Italian lexicon as a whole (see above). Finally, the stress neighborhoods of the words (operationally defined as in Spinelli, Sulpizio et al. (2017) as the set of words that share the last three letters) included a mean of 598.31 dominant-stress words (the *dominant-stress-neighborhood count*), corresponding to a mean percentage of 73% of the stress neighborhood (the *dominant-stress-neighborhood proportion*).

Next in the process of creating the nonword stimuli, 1–4 of the source word's letters were replaced with other letters (excluding the last 3 letters and excluding letter sequences corresponding to Italian affixes, regardless of whether the letter sequence was in fact an affix in the word). The substitution letters were selected based on their frequency in the language (i.e., letters occurring more frequently in the words in the database were more likely to be selected than letters occurring less frequently) with the further constraints that, first, consonants were replaced with consonants and vowels with vowels; second, the result of the replacements was not an existing Italian word; and third, the resulting nonword was phonotactically legal (e.g., for the source word albero ' tree', the nonword "acbero" would not be acceptable since "cb" is an illegal consonant cluster in Italian). The replaced letter proportion (i.e., the number of replaced letters divided by the total number of letters) was 33% on average. In replacing the letters, we made certain that the replaced-letter proportion would not be correlated with length, $r = -.025$, $p = .473$. Further, because of the constraints we used when replacing the letters and because of the fact that the substitution letters were selected using a frequency-weighted procedure, the bigram frequency of the nonwords we created was similar, albeit slightly lower, than that of their source words (M = 11.66 and M = 11.70, respectively) and, most importantly, it did not correlate with replaced-letter proportion either, $r = -.044$, $p = .218$.

A number of additional variables was measured for the nonwords created (see Table 4). Finally, for the nonwords in which a suffix was present (673), we also determined whether that suffix was uniquely associated with the dominant (437 nonwords) or nondominant stress patterns (99 nonwords), or whether that suffix could be associated with both dominant or nondominant stress patterns (137 nonwords).

Table 4

Characteristics of the nonwords used in the megastudy

Characteristic	*M*	*SD*	Range
Replaced letter proportion	.33	0.13	.09–.57
Length in letters	7.79	1.79	5–13
Bigram frequency	11.66	0.32	10.63–12.49
Length in syllables	3.52	0.67	3–5
N (orthographic neighborhood size)	1.40	2.32	0–18
Mean *N* frequency (occurrences per million)	6.84	24.88	0–403.63
Dominant-stress-neighborhood count	598.31	721.65	2–2,870
Dominant-stress-neighborhood proportion	.73	0.35	.04–1
Penultimate syllable stress count	665.23	703.95	0–2,854
Penultimate syllable stress proportion	.71	0.24	0–1
Antepenultimate syllable stress count	117.14	110.52	0–461
Antepenultimate syllable stress proportion	.20	0.16	0–1

Procedure

Participants were tested individually in a quiet room. They were required to sit in front of a computer screen and to wear a headset with a microphone through which their responses would be recorded. Each trial started with a fixation cross (“+”) displayed in the center of the screen for 250 ms, followed by the nonword, which was presented in the same position in lowercase 12-pt Courier New font and remained on the screen for 2500 ms, regardless of the participant’s response time. All stimuli were presented in black against a white background. Participants were instructed to read the nonwords aloud as quickly and as accurately as possible, the standard instructions used in most reading aloud experiments. We used those instructions to reduce the possibility of participants engaging in strategic behaviors which might override the response which would come more natural for them to produce in a speeded task. On the other hand, the fact that somewhat long nonwords were included raised the possibility of participants starting articulating the nonword before they had completely processed it and had decided where to place stress. To address this issue, the experimenter emphasized that the pronunciation should be, as much as possible, natural and without hesitations and that it should be started only once the whole nonword had been processed. The experiment started with 10 practice nonwords, followed by the list of 800 experimental nonwords arranged in four blocks of 200 nonwords each. The list was the same for all participants and the order of presentation of the nonwords was randomized for each participant. Participants were allowed to take a break at the end of each block. The whole session lasted approximately 45 minutes, including breaks. Stimulus presentation and data recording were controlled with DMDX software (Forster & Forster, 2003).

Data Analysis

Response recordings and waveforms were manually inspected with CheckVocal (Protopapas, 2007) by one of the authors (S. T., a native Italian speaker) in order to categorize the stress pattern assigned by participants and to determine the accuracy of the response (using the response recording), and to ensure the correct placement of the response onset for RT measurement (using the response waveform). Some studies showed that stressed syllables tend to have longer vowel duration, higher pitch, and greater intensity than unstressed syllables (Cutler, 2005; Lehiste, 1970). In Italian, stress is mainly reflected in duration differences between syllables with primary stress and unstressed syllables. For syllables in non-final position, these differences are especially pronounced compared to other languages (Alfano et al., 2009; Bertinetto, 1980; Eriksson et al., 2016). As a result, there is typically high consistency among judges categorizing stress responses in Italian (Sulpizio et al., 2013).

However, to ensure that the categorization provided by S. T. was indeed accurate, G. S. and L. C. (who are also native Italian speakers) conducted an acoustic analysis using Praat, version 6.3.10 (Boersma & Weenink, 2023) software for a random subset of 1800 observations, i.e., 5% of the total observations, selected with the constraint that they should not contain missing responses or errors (see below) and roughly split between G. S. and L. C. This analysis involved manually labelling the acoustic boundaries of the vowel corresponding to the nucleus of each syllable in the nonword and then extracting its duration, intensity, and pitch. For intensity and pitch, the maximum values were extracted using Praat's default settings (Olivucci et al., 2016). (Pitch could not be extracted for 46 vowels out of the 6293 processed ones, likely because of an almost voiceless articulation of those vowels.)

Considering the nonwords categorized by S. T. as being stressed on either the penultimate position (1424 observations), the antepenultimate position (373 observations), or the pre-antepenultimate position (2 observations) and the duration, pitch, and intensity values for those syllables, the vowel of the syllable categorized as being stressed was longer than that of the other two in 94.22% of cases, had a higher pitch in 77.99% of cases (excluding those for which pitch for either of the three syllables could not be calculated), and had a greater intensity in 32.52% of cases. (Note 5) Breaking down those cases by the stressed syllable's position revealed that the vowel of the syllable categorized as being stressed was typically longer than that in the other two regardless of its position in the nonword (penultimate: 94.24%; antepenultimate: 94.10%; pre-antepenultimate: 100%). The vowel of the syllable categorized as being stressed typically had a higher pitch only in penultimate position (87.84%;

antepenultimate: 41.13%; pre-antepenultimate: 0%) and had greater intensity only in antepenultimate position (61.93%; penultimate: 24.79%; pre-antepenultimate: 50%).

Overall, S. T.'s stress judgments largely aligned with the relative duration of the relevant vowels, consistent with the idea that duration is the strongest and most reliable indicator of lexical stress in Italian (Alfano et al., 2009; Bertinetto, 1980; Eriksson et al., 2016). For pitch and intensity, the situation was more complex. As a further attempt to establish the accuracy of S. T.'s judgments, G. S. categorized the stress pattern of all 1800 observations independently. G. S. and S. T.'s judgments agreed on 99.72% of cases, $\kappa = .992$, $p < .001$, i.e., all cases except 4. Those 4 cases (all of which being among those cases for which either duration, pitch, or intensity differences did not support S. T.'s categorization) were reviewed by L. C. and S. T., who agreed with G. S.'s judgment. The stress patterns for the affected observations were corrected accordingly. However, overall, the results of the acoustic analysis and the high agreement between the two judges suggested that S. T.'s categorization of stress patterns was very accurate. Therefore, we relied on that categorization for subsequent analyses. Further, these data suggest that the evaluation of stress position is much easier in Italian than, for example, in English (e.g., Mousikou et al., 2017).

Missing responses were discarded from the analyses (15 trials, corresponding to 0.04% of the data). Responses that involved any error in pronunciation were phonetically transcribed and were discarded from the analyses (1740 trials, corresponding to 4.83% of the data). These mispronunciations included substitutions, additions, deletions or transpositions of one or more phonemes or syllables, responses decomposed in two or more parts (e.g. [dira] [tente] for the nonword "diratente"), false starts, or the existence of more than one error for the same stimulus. (Note 6) In the remaining responses, there were no RTs below 300 ms, therefore no additional responses were removed.

For each response, the stress pattern assigned to the nonword was categorized as ultimate (3 trials), penultimate (26,450 trials), antepenultimate (7,770 trials), or preantepenultimate (22 trials). Given the anticipated (see Colombo, 1992; Colombo & Zevin, 2009; Colombo et al.,

2014) result that the penultimate responses were the most frequent (77.24%), those responses were then categorized as dominant responses and the other responses (mainly, antepenultimate responses) were categorized as nondominant responses. Prior to the analysis, nonwords with a heavy penultimate syllable were inspected. They received dominant stress in the vast majority (99.28%) of the observations, a prevalence of dominant stress than was even stronger than the already strong prevalence observed with words in the corpus analysis (95.83%). Those nonwords (accounting for 6,397 observations) were treated as fillers and not included in the analysis, which was based on 27,848 observations in total.

Stress responses (dominant vs. nondominant) were analysed in the same way as stress patterns were analyzed in the corpus analysis, except that generalized linear *mixed-effects* modelling was used, implemented with the glmer function in the lme4 package, version 1.1-23 (Bates et al., 2015). This type of modelling allowed us to specify, in addition to the predictors or fixed effects, also random effects on stress assignment performance, i.e., participants and items (i.e., the nonwords). The models were fit by maximum likelihood with the Laplace approximation technique. Because in the current version of lme4, convergence failures for generalized linear mixed-effects models are frequent (although many of those failures reflect false positives: Bolker, 2022), in order to limit the occurrence of convergence failures, we kept the random structure of the models as simple as possible by using only random intercepts for participants and items and by standardizing all continuous predictors. For the same reason, the number of maximum iterations for model evaluation was raised to one million and model evaluation was conducted with the BOBYQA optimizer, an optimizer that typically returns estimates that are equivalent to those returned by lme4's default optimizer but that results in fewer convergence failures (see, e.g., Colombo et al., 2020; Lupker et al., 2020a, 2020b). When convergence failures did occur, we followed the recommended troubleshooting procedures (see the "convergence" help page in R), including restarting the fit from the apparent optimum and using other optimizers to ensure the convergence warning was a false positive.

For the fixed effects, we used a hierarchical approach similar to that used in the corpus analysis. Specifically, predictors of stress responses were entered in 5 steps. In Step 0, the *null model*, we entered the random effects only. In Step 1, we added the *stress-neighborhood predictors*: dominant-stress count and proportion in the stress neighborhood. In Step 2, we added the *syllabic predictors*: length in number of syllables and penultimate and antepenultimate syllable stress count and proportion. In Step 3, we added the *morphological predictors*: dominant-stress suffix and nondominant-stress suffix, both dummy-coded as present vs. absent. Finally, in Step 4, we added the *lexical predictors*, that is, the predictors relevant to the source word: source-word stress (dominant vs. nondominant), source-word log frequency, replaced-letter proportion (i.e., the proportion of letters replaced in the source word to create the nonword), as well as the two-way interactions between replaced-letter proportion and source-word stress, on one hand, and between replaced-letter proportion and source-word log frequency on the other.

As noted above, our expectation for the sublexical predictors, entered in Steps 1–3, was for those predictors to have similar impacts as they did in the corpus analysis (with the potential exception of the penultimate- and antepenultimate-syllable stress counts, which did not have clear impacts in that analysis and were eventually excluded). For the lexical predictors entered in Step 4, we expected a tendency for nonwords to receive the same stress pattern as their source words, especially with proportionally fewer replaced letters (because, for those stimuli, the nonword resembled the source word to a greater extent) and/or with more frequent source words (because, for those stimuli, the source word would be more easily identified).

Model building followed the same procedure as that used in the corpus analysis. That is, for each step (excluding Step 0), we dropped the fixed effects that did not reach statistical significance and we compared the resulting model with the final model from the previous step with a chi square test to determine whether goodness of fit in the current step (i.e., the more complex model) had improved from the previous step (i.e., the less complex model). Estimates of, in this case, R_F^2 (i.e., the proportion of total variance explained by the *fixed* effects, hence the “F” in the subscript; Zhang, 2017, 2022a) were also obtained for the final model in each step.

Also similar to the corpus analysis, the fixed effects were inspected for multicollinearity before proceeding with the analyses. We ran a model in which the random effects and all fixed effects (excluding interactions between them) were entered simultaneously and, on the basis of this model, we calculated the variance inflation factor. This value was below 3 for all fixed effects, posing little concern for multicollinearity. Furthermore, all the correlation coefficients for the (standardized) continuous variables used as fixed effects in the stress and latency analyses were less than .75 (Cohen et al., 2003). Therefore, no further adjustment for multicollinearity seemed to be required. (Note 7)

Results and Discussion

The basic relationships between each individual predictor and the mean proportion of dominant-stress responses for each of the nonwords used in this analysis are reported in Table 5 in the form of Pearson's correlations (for continuous predictors) and means (for categorical predictors). Further, for each predictor and for the interactions between the predictors that we examined, a plot of the mean proportion of dominant-stress responses for each of the nonwords used in this analysis is displayed in Figure 4 for the stress-neighborhood predictors, in Figure 5 for the syllabic predictors, in Figure 6 for the morphological predictors, and in Figure 7 for the lexical predictors. In each plot is also displayed the relevant trend line (for continuous predictors) or marginal mean (for categorical predictors) along with its 95% confidence interval. These values were estimated with a model in which all predictors were entered simultaneously.

Examination of the correlation values and graphs suggests a noticeable impact on stress assignment for both stress neighborhood count and proportion among the stress-neighborhood predictors, penultimate-syllable stress proportion among the syllabic predictors, dominant-stress and especially nondominant-stress suffix presence among the morphological predictors, and source-word stress and its interaction with replaced-letter proportion among the lexical predictors.

Table 5

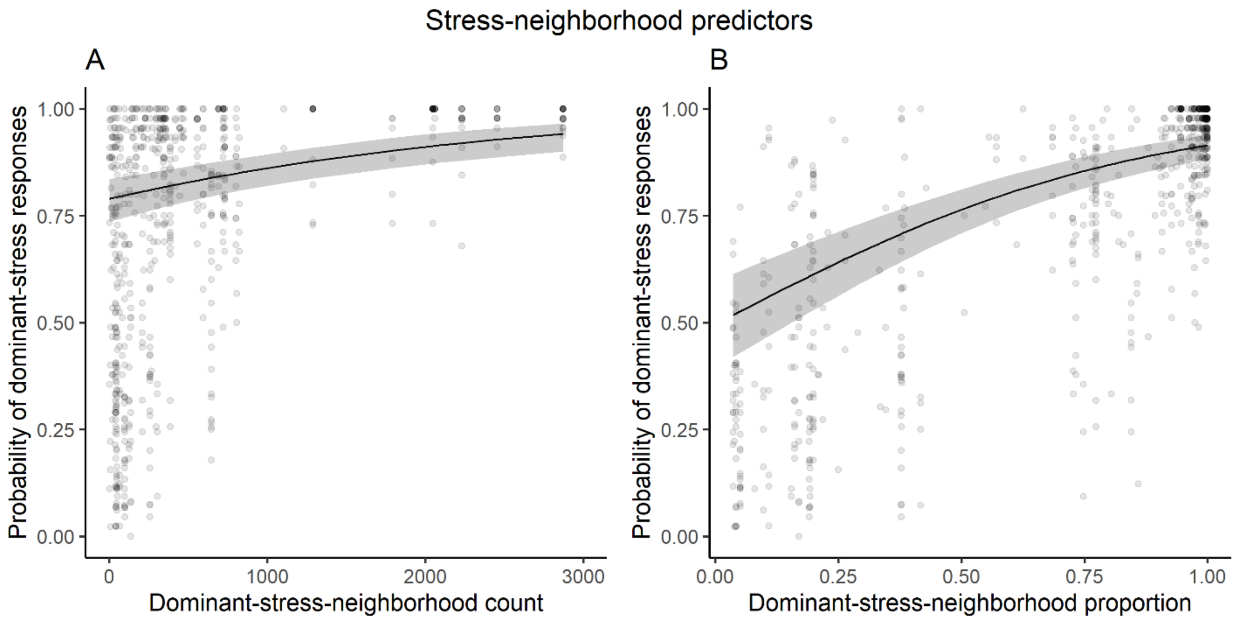
The basic relationships between predictors and proportion of dominant-stress responses to the nonwords used in the megastudy

Type of predictor	Predictor	Association with stress responses
Stress-neighborhood	Dominant-stress-neighborhood count	r = .418
	Dominant-stress-neighborhood proportion	r = .773
Syllabic	Length in syllables	r = −.034
	Penultimate-syllable stress count	r = .237
	Penultimate-syllable stress proportion	r = .520
	Antepenultimate-syllable stress count	r = −.006
	Antepenultimate-syllable stress proportion	r = −.048
Morphological	Dominant-stress suffix presence	Absent: M = .55 Present: M = .80
	Nondominant-stress suffix presence	Absent: M = .83 Present: M = .52
Lexical	Source-word stress	Dominant: M = .87 Nondominant: M = .47
	Replaced-letter proportion	r = .065
	Source-word log frequency	r = .004

Note. The critical value for significance (.05, two-tailed) for the correlations is r_c = .077.

Figure 4

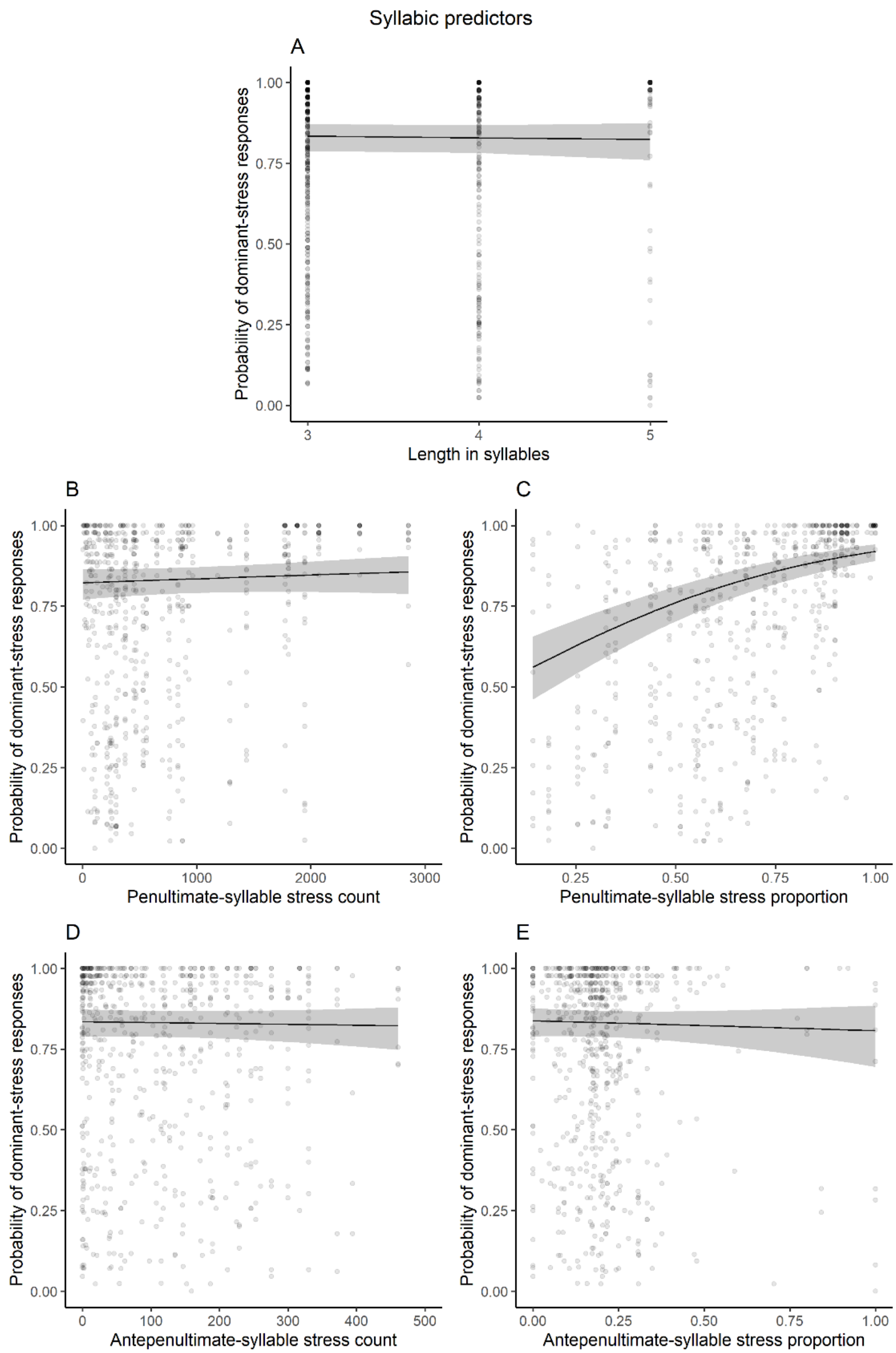
The impact of the stress-neighborhood predictors on the proportion of dominant-stress responses to the nonwords used in the megastudy



Note. The lines represent estimated intercepts and trends for the probability of dominant stress patterns assigned as a function of dominant-stress-neighborhood count (A) and proportion (B).

Figure 5

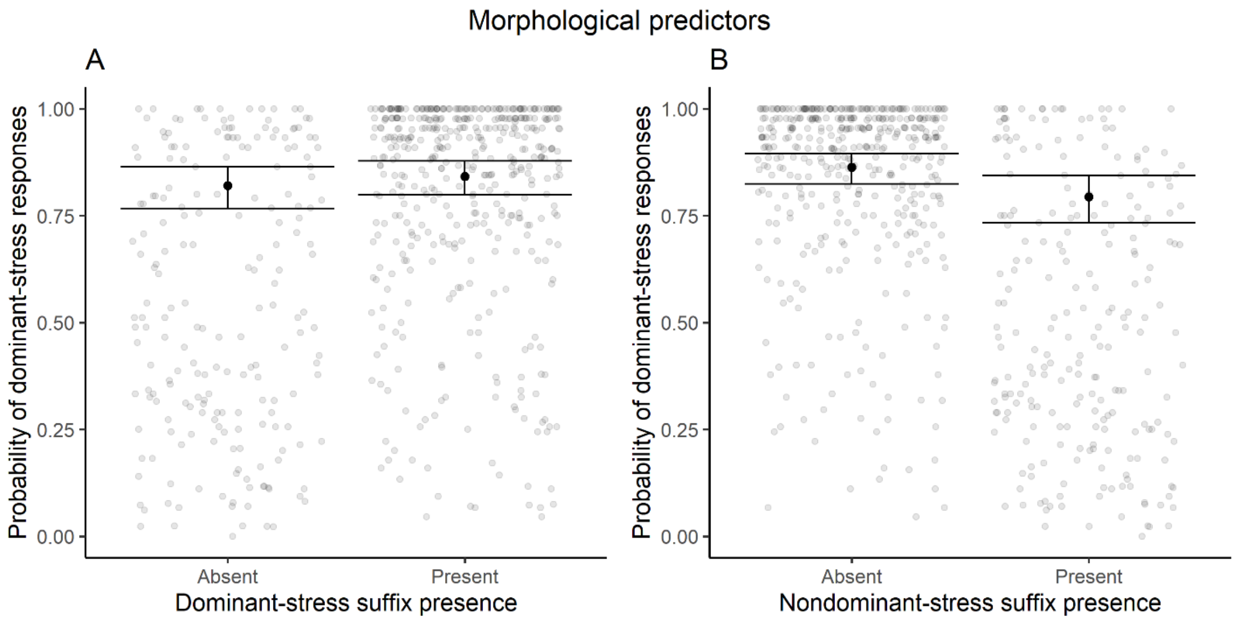
The impact of the syllabic predictors on the proportion of dominant-stress responses to the nonwords used in the megastudy



Note. The lines represent estimated intercepts and trends for the probability of dominant stress patterns assigned as a function of length in syllables (A), penultimate-syllable stress count (B) and proportion (C), and antepenultimate-syllable stress count (D) and proportion (E).

Figure 6

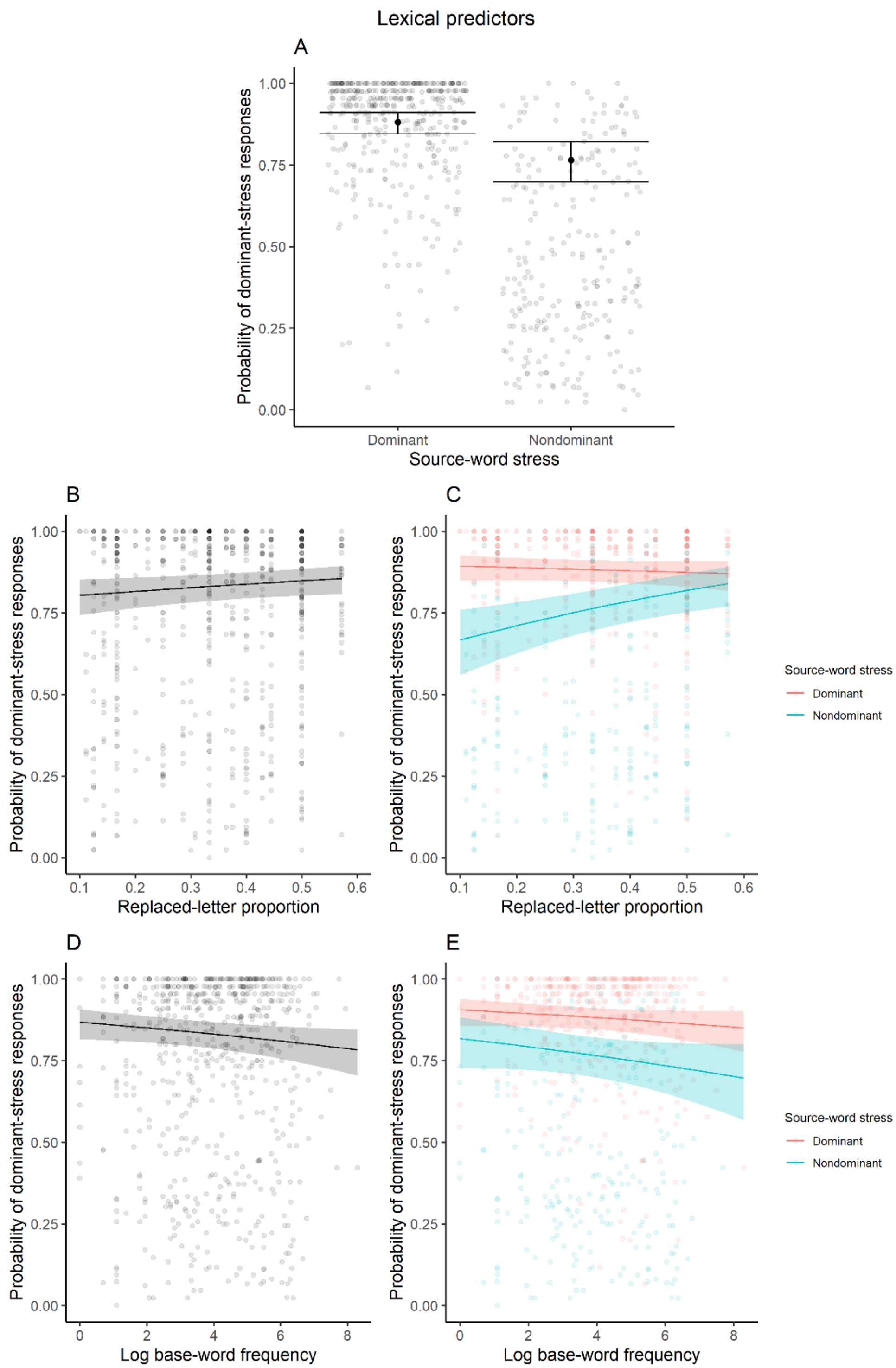
The impact of the morphological predictors on the proportion of dominant-stress responses to the nonwords used in the megastudy



Note. The darker circles represent the estimated marginal means for the probability of dominant stress patterns assigned as a function of the presence of a dominant-stress suffix (A) and nondominant-stress suffix (B).

Figure 7 [Requires colors in the online version]

The impact of the lexical predictors on the proportion of dominant-stress responses to the nonwords used in the megastudy



Note. The darker circles in the top graph (A) represent the estimated marginal means for the probability of dominant stress patterns assigned as a function of the source-word stress. In the other graphs, the lines represent estimated intercepts and trends for the probability of dominant stress patterns assigned as a function of: replaced-letter proportion examined overall (B); the interaction between source-word stress and replaced-letter proportion (C); log source-word frequency examined overall (D); the interaction between log source-word frequency and source-word stress (E). In graphs C and E, individual means and regression lines for nonwords with dominant and nondominant-stress source words are marked with red and aqua blue, respectively (to view colors, see the online version of this paper).

The hierarchical regression results are in general agreement with this assessment. These results are reported in Table 6. In that table, we report, for each step, the fixed effects retained in the final model for that step, along with the parameters from the very final model in Step 4 (i.e., the best-fitting model). For each step, we also report the R_F^2 and the chi square test comparing the model in that step with the model from the previous step. Here, we describe the main patterns of results.

Table 6

Hierarchical regression results for stress responses in the megastudy

Step	Fixed effects	R_F^2	Model comparison		Best-fitting model parameters			
			χ²	*p*	β	*SE*	*z*	*p*
Step 1: Stress-neighborhood predictors		.2514	560.36	< .001				
	Dominant-stress-neighborhood count				0.394	0.072	5.49	< .001
	Dominant-stress-neighborhood proportion				0.875	0.091	9.60	< .001
Step 2: Syllabic predictors		.2766	103.56	< .001				
	Penultimate-syllable stress proportion				0.599	0.060	9.99	< .001
Step 3: Morphological predictors		.2786	7.91	.005				
	Nondominant-stress suffix				0.236	0.070	3.35	< .001
Step 4: Lexical predictors		.2924	42.61	< .001				
	Source-word stress				0.426	0.078	5.44	< .001
	Replaced-letter proportion				0.105	0.055	1.92	.055
	Source-word stress × Replaced-letter				−0.160	0.055	−2.91	.004
	Log source-word frequency				−0.113	0.054	−2.08	.037

Note. For each step, the model comparison is always between the model in that step and the model in the previous step. For the Step-1 model, the comparison is with the Step-0 model, i.e., the null model including random effects only. Note that all continuous fixed effects were standardized prior to model evaluation.

First, for the *stress-neighborhood predictors* examined in Step 1, we confirmed a role for dominant-stress-neighborhood count and proportion: Nonwords with an ending associated with a higher number and proportion of dominant-stress words in the lexicon were more likely to receive dominant stress (see Figure 4). A few things should be noted about these patterns: First, the impact of the proportion of stress neighbors (Figure 4B) is stronger than that of the count (Figure 4A), as can be seen also by a comparison of the r and z values (Tables 5 and 6). Second, Figure 4B shows that when the proportion of dominant stress neighbors is 0 or very low, the estimated probability of dominant stress responses is the lowest, showing the influence of neighbors with antepenultimate stress. Still, the intercept level is around 50%, as there are quite a few dominant stress responses even when the stress neighborhood does not support dominant stress. Further, when stress neighborhood is around 50% and thus essentially neutral, the estimated probability of dominant stress responses increases to around 75% (although note that there are relatively few nonwords with a dominant-stress-neighborhood proportion around 50%, reflecting the general situation in Italian in which this type of ambiguous stress neighborhood is uncommon). Thus, even though stress-neighborhood predictors explain a large portion of the data (over 25%, see Table 6), there also appears to be a general tendency for assigning the dominant stress pattern, consistent with the idea of stress dominance playing a role.

Second, for the *syllabic predictors* examined in Step 2, we found an effect of the proportion of words having the same syllable stressed in the penultimate syllable position. This effect indicated that, above and beyond the other effects included in the model, dominant-stress assignment was more likely for nonwords with a penultimate syllable that, in the lexicon, typically attracts stress (i.e., it is more often stressed than unstressed; see Figure 5C). In contrast, we found no significant effects for the other syllable-related predictors (i.e., neither the count for the penultimate syllable nor the count and proportion for the antepenultimate syllable). Length in syllables had no specific effect on the proportion of dominant stress assigned, presumably due to the widespread proportion of penultimate stress over all stimulus lengths.

Third, for the *morphological predictors* examined in Step 3, we only found an effect of the nondominant-stress suffix. This effect indicated, as expected, that for nonwords with a nondominant-stress suffix, compared to nonwords without it, there was a significant decrease in dominant-stress responses (from 83% to 52% in the observed data; see Figure 6B). Note also, however, that for nonwords with a dominant-stress suffix the percentage of dominant stress responses was numerically higher than for nonwords without that type of suffix (from 80% to 55% in the observed data, see Table 5 and Figure 6A), an effect that was not significant in the model. This nonsignificant increase may be partially due to correlated predictors (e.g., the stress-neighborhood predictors) explaining that portion of variance. However, it may also reflect the fact that participants were biased to assign the dominant stress even in the absence of a dominant suffix. This bias can be appreciated best by examining "ambiguous" suffixes, that is, suffixes, such as "-ano", which are associated with both dominant and nondominant stress in Italian: Despite the suffix being ambiguous, nonwords with this type of suffix were assigned dominant stress 67% of the time. (Note 8)

Note, in any case, that when a suffix was present, either a dominant-stress one or a nondominant-stress one, the stress patterns assigned were not always consistent with that suffix (e.g., when a nondominant-stress suffix was present, there were still 52% dominant-stress responses, as noted, a proportion which dropped but remained quite high, i.e., 33%, when excluding ambiguous suffixes). Therefore, it would seem unlikely that recognition of a suffix resulted in a rule being applied as proposed in rule-based models such as Rastle and Coltheart's (2000). Such a pattern would seem to be more likely to be captured by models based on probabilistic associations (see Study 3 below).

Finally, for the *lexical predictors* examined in Step 4, we found, first, a main effect of source-word stress indicating that, in general, dominant stress was preferred for nonwords with a dominant-stress source word (87% in the observed means) whereas for nonwords with a nondominant-stress source-word, dominant-stress responses were approximately as likely as nondominant-stress responses (47% and 53%, respectively; see Figure 7A). We also found a marginal main effect of replaced-letter proportion indicating that, in general, nonwords with a higher proportion of replaced letters (i.e., nonwords that were less similar to their source words) tended to receive dominant stress (see Figure 7B).

Most importantly, there was also an interaction between source-word stress and replaced-letter proportion. This interaction is represented in Figure 7C. What this interaction indicates is that the effect of replaced-letter proportion depended on the source-word stress pattern. Specifically, when the source word had a dominant stress pattern, replaced-letter proportion had essentially no effect, $\beta = -0.055$, $SE = 0.071$, $z = -0.77$, $p = .440$ (see the red line in the figure). In contrast, when the source word had a nondominant stress pattern, there was a significant effect of replaced-letter proportion, $\beta = 0.265$, $SE = 0.084$, $z = 3.18$, $p = .002$, with dominant stress being more likely with a higher proportion of replaced letters (i.e., when the nonword was less similar to its nondominant-stress source word; see the aqua blue line in the figure). Essentially, the effect of source-word stress was reasonably strong when the replaced-letter proportion was low (i.e., when the nonword was more similar to its source word). In contrast, when the replaced-letter proportion was high (i.e., when the nonword was less similar to its source word), the effect of source-word stress was reduced because participants tended to assign the dominant stress to all nonwords with little regard to the stress pattern of their source word.

Finally, there was also a small main effect of source-word log frequency, with nonwords with a more frequent source word being less likely to receive dominant stress (see Figure 7D). As anticipated, this tendency was mainly driven by nonwords with a nondominant stress source-word (see the aqua blue line in Figure 7E), suggesting that assigning dominant stress was less likely when the nonword activated a frequent nondominant-stress word. However, the interaction between source-word stress and log source-word frequency did not reach significance.

Study 3: Comparison with computational models of reading

After the examination of the cues to stress in the word corpus and in the nonword megastudy, what we would like to address is how the two computational models of polysyllabic word reading for Italian, Pagliuca and Monaghan's (2010) model and CDP++ Italian (Perry et al., 2014), which we will now review in greater detail than in the Introduction, would be able to explain the present results.

Pagliuca and Monaghan's (2010) model is a parallel distributed processing (PDP) model of reading in Italian, with the standard three-layer structure of such models: orthographic, hidden, and phonological (Harm & Seidenberg, 1999, 2004). Different from most PDP models, however, the model is able to process words ranging from one and three syllables in length and to process stress information. In Pagliuca and Monaghan's (2010) simulations, the model was fairly successful at reading words and nonwords and at reproducing relevant empirical findings such as frequency and morphological effects.

With respect to stress, the authors conducted one simulation using a sample of word stimuli from Burani and Arduino's (2004) Experiment 1, a reading aloud experiment involving dominant and nondominant stress words with stress neighborhoods consistent vs. inconsistent with the words' stress pattern. The original experiment produced only a main effect of consistency with the stress neighborhood (with words with a stress pattern consistent with their stress neighborhood being faster than words with a stress pattern inconsistent with their stress neighborhood) but no main effect of stress dominance nor an interaction. In contrast, in the model's simulation, effects only emerged in the stress errors, which were frequent (27%), and the result pattern was somewhat different: Neither main effect emerged whereas the interaction did, with nondominant stress words producing fewer errors when their stress neighborhood was consistent vs. inconsistent with that pattern. That is, the model simulated a regular effect of consistency with the stress neighborhood for nondominant-stress words but no such effect for dominant-stress words. As we have noted in the Introduction, however, Colombo (1992) did report this type of pattern in a previous experiment (Experiment 4) with a similar design as Burani and Arduino's (2004) Experiment 1 but using different stimuli.

How Pagliuca and Monaghan's (2010) model would fare at assigning stress to nonwords, the type of stimuli used in the present megastudy, is an open question and one that, unfortunately, we cannot directly answer through simulation because the programming language used for the model is, at present, no longer actively maintained. What may help in answering that question, however, is noting that the focus of the model was to demonstrate that even in a language such as Italian in which there is a one-to-one correspondence between letters and phonemes for most letters, a single-route parallel distributed processing model can pick up effects that seem to emerge from units of larger grain size than the size that would appear sufficient for reading in such a language (i.e., one letter). Indeed, the model was able to simulate morphological effects for nonwords, such as the finding that nonwords which include a true root and a true suffix but which do not exist in the language (e.g., “donnista”, with “donn-“ being a true root and “-ista” being a true suffix) are read faster than control nonwords with include neither a true root nor a true suffix (e.g., “dennosto”; Burani et al., 2008). Note that the model had no access to morphological or semantic information. Thus, the fact that the model was able to simulate this particular result suggests that the network learned to use units of relatively large grain size such as roots and suffixes. Thus, in principle, the model would seem to be able to use large grain size units that pertain to stress as well, including endings, syllables, and suffixes, and therefore, potentially, to simulate the impact of the sublexical cues that our megastudy revealed. That said, the model’s partial failure to reproduce Burani and Arduino’s (2004) pattern suggests that some change is required in the model. Further, it is not quite clear how the model would produce the overall bias for dominant stress revealed by our megastudy, as that bias appears to be tied to no particular unit of any grain size.

The other computational model of polysyllabic word reading for Italian is the Italian version of the Connectionist Dual-Process model of reading aloud, CDP++ Italian (Perry et al., 2014). As with earlier versions of the CDP model, CDP++ Italian involves a lexical route connecting an orthographic and a phonological lexicon, and a sublexical route consisting of a graphemic parser and a two-layer associative (TLA) network connecting graphemes and phonemes trained on a large Italian corpus (although not the typical one used in Italian psycholinguistic studies). The two routes converge onto a phonemic buffer where the outputs from both routes are integrated, with outputs including both segmental information (i.e., phonemes) and suprasegmental information (i.e., stress patterns).

As Pagliuca and Monaghan's (2010) model, CPD++ Italian was shown to simulate several empirical findings of relevance. Of interest here are the simulations involving stress, particularly those for Burani and Arduino's (2004) Experiment 1 and Colombo's (1992) Experiment 4. As noted, Burani and Arduino's (2004) reading aloud experiment produced an effect of consistency with the stress neighborhood for dominant- and nondominant-stress words alike. In contrast, Colombo (1992) only found an effect of consistency with stress neighborhood for nondominant-stress words. CDP++ Italian was able to simulate both patterns of results for the respective stimuli. The model was also able to closely reproduce results for stress assignment in nonword reading (Colombo, 1992, Experiment 5; Colombo & Zevin, 2009). On the other hand, the model was not able to reproduce the main effect of stress dominance reported by Colombo (1992) for her Experiment 4, nor the effect of the number of stress friends (i.e., the words in the stress neighborhood consistent with the word's stress pattern) for Burani and Arduino's (2004) Experiment 2. Further, it tended to overpredict the effect of stress neighborhood for Sulpizio et al.'s (2013) stimuli. In general, in any case, CDP++ Italian was reasonably successful at simulating empirical findings relevant to lexical stress, although it provided no explanation for the inconsistencies among some of them, particularly that between Burani and Arduino's (2004) Experiment 1 and Colombo's (1992) Experiment 4.

How would CDP++ Italian fare at simulating the pattern of results produced by our megastudy? Concerning the stress-neighborhood effects, because CDP++ Italian was able to simulate those effects with nonwords before, that model could be presumed to have no difficulty at doing so with our stimuli. In contrast, it is not clear that CDP++ Italian would be able to simulate the other effects the present megastudy produced, some of which are novel (e.g., the syllabic, morphological, and lexical ones) and some of which are effects with which CDP++ Italian has had some difficulty before (e.g., stress dominance and count measures of stress neighborhood).

Here, we were able to provide a direct answer to this question by presenting CDP++ Italian with some of the nonwords used in the megastudy. Not all of the nonwords could be used because of CDP++ Italian's inability to process stimuli with more than 8 letters or of more than 3 syllables in length. The phonological representations produced by the model for the remaining nonwords (424) were then contrasted with (manually coded) acceptable phonological representations for those nonwords, with mismatches being coded as inaccurate (as noted in footnote 6, in Italian, a nonword typically only has one acceptable representation at the segmental level). There were 48 nonwords for which the model gave an inaccurate response, of which most (43) were lexicalizations (i.e., words contained in CDP++'s lexicon). We excluded those nonwords as well and conducted generalized linear model analyses (the same type of analysis used for the corpus analysis) of the model latencies (i.e., processing cycles) for all of the remaining nonwords (376) and of the stress responses for the nonwords with a light penultimate syllable (314). We focus on the stress responses here (for the latencies, see the Supplementary Materials).

Before presenting those results, however, for comparison purposes, we present in Table 7 the results of a re-analysis of the stress responses in our human data based on only the nonwords that CDP++ Italian could process and for which it produced accurate responses. Essentially, the pattern of results was the same as in the analysis with all nonwords with the only exception being that there was no effect of nondominant-stress suffix presence here (i.e., no morphological predictor was significant; compare Tables 6 and 7), perhaps due to the reduced power afforded by this analysis (i.e., there were only 38 nonwords with a suffix uniquely associated with nondominant stress in this reduced set of nonwords, down from 99 in the full set).

Table 7

Hierarchical regression results for stress responses in the megastudy using nonwords compatible with CDP++ Italian

		R_F^2	Model comparison		Best-fitting model parameters			
Step	Fixed effects		χ^2	*p*	β	*SE*	*z*	*p*
Step 1: Stress-neighborhood predictors		.2161	248.37	< .001				
	Dominant-stress-neighborhood count				0.338	0.097	3.48	< .001
	Dominant-stress-neighborhood proportion				1.12	0.110	10.13	< .001
Step 2: Syllabic predictors		.2278	26.26	< .001				
	Penultimate-syllable stress proportion				0.396	0.082	4.85	< .001
Step 3: Morphological predictors		.2300	4.28	.117				
Step 4: Lexical predictors		.2417	21.03	< .001				
	Source-word stress				0.253	0.112	2.25	.024
	Replaced-letter proportion				0.156	0.079	1.99	.047
	Source-word stress × Replaced-letter				−0.260	0.079	−3.30	< .001
	Log source-word frequency				−0.157	0.078	−2.03	.043

Note. For each step, the model comparison is always between the model in that step and the model in the previous step that was not discarded (i.e., the Step-4 model was compared with the Step-2 model because the Step-3 model was discarded). For the Step-1 model, the comparison is with the Step-0 model, i.e., the null model including random effects only. Note that all continuous fixed effects were standardized prior to model evaluation.

The results of the analysis of CDP++ Italian's stress responses are reported in Table 8. Although the overall proportion of dominant-stress responses in the simulation (72.29%) was similar to that of the human data for these nonwords (74.46%), as can be seen from Table 8, the final model for the simulation included only two predictors: dominant-stress-neighborhood proportion (with nonwords with an ending associated with a higher proportion of dominant-stress words in the lexicon being more likely to receive dominant stress) and dominant-stress suffix presence (with nonwords with a dominant-stress suffix being more likely to receive dominant stress). Thus, CDP++ Italian failed to simulate the impact of many cues Italian readers use to assign stress to nonwords, including dominant-stress-neighborhood count, penultimate-syllable proportion, and all of the lexical cues.

Table 8

Hierarchical regression results for stress responses in the simulation with CDP++ Italian

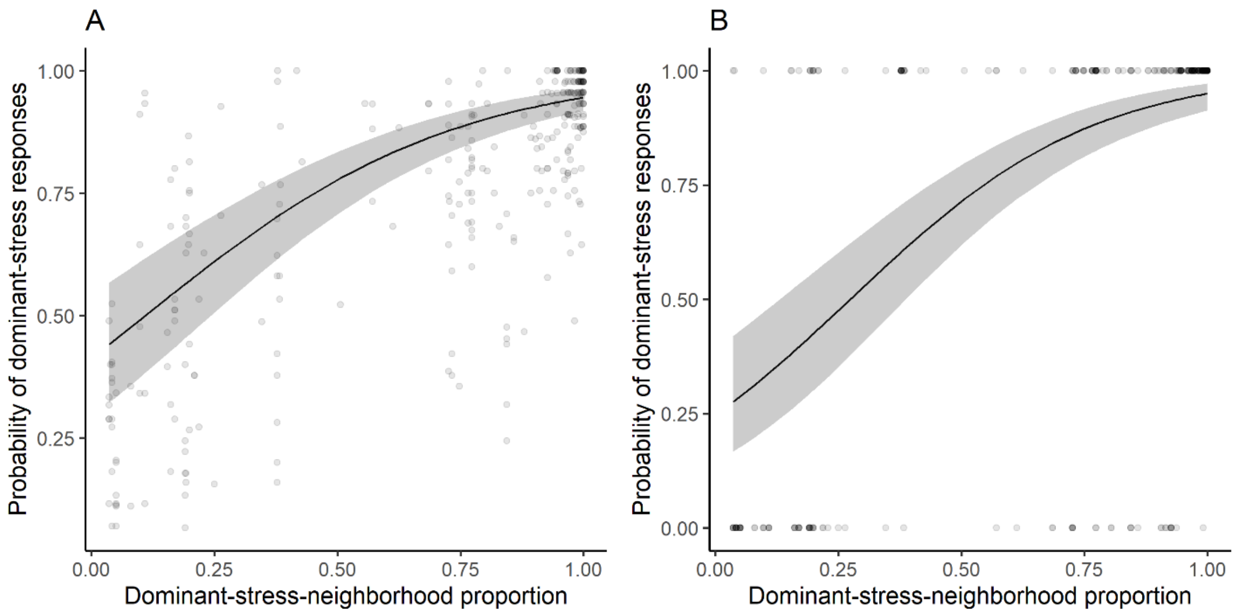
		R^2	Model comparison		Best-fitting model parameters			
Step	Fixed effects		χ^2	*p*	β	*SE*	*z*	*p*
Step 1: Stress-neighborhood predictors		.3712	116.6	< .001				
	Dominant-stress-neighborhood proportion				1.42	0.164	8.66	< .001
Step 2: Syllabic predictors		.3758	3.59	.464				
Step 3: Morphological predictors		.3936	8.47	.004				
	Dominant-stress-suffix				−0.48	0.163	−2.91	.004
Step 4: Lexical predictors		.4042	4.09	.537				

Note. For each step, the model comparison is always between the model in that step and the model in the previous step that was not discarded (i.e., the Step-3 model was compared with the Step-1 model because the Step-2 model was discarded). For the Step-1 model, the comparison is with the Step-0 model, i.e., the null model including the intercept only. Note that all continuous predictors were standardized prior to model evaluation.

Upon closer inspection, it is worth noting that CDP++ Italian seemed to underestimate the proportion of dominant stress responses when the dominant-stress-neighborhood proportion is low. This fact can be appreciated by comparing the impact of this predictor in the final models for the human data (Figure 8A) and the CDP++ Italian simulation data (Figure 8B). In particular, at the intercept (i.e., when dominant-stress-neighborhood proportion is close to zero), dominant stress is still almost as probable as nondominant stress in the human data (with the lowest estimated probability being 44%) whereas dominant stress probability is quite a bit lower in the CDP++ Italian simulation data (with the lowest estimated probability being 28%). On the other hand, when stress neighborhood is around 50% and thus essentially neutral, dominant stress probability is similar in the human (78%) and simulation data (72%), with both favoring dominant stress. In other words, CDP++ Italian seems to be more influenced by stress neighborhood than humans. This fact can also be appreciated by comparing the variance accounted for by the stress neighborhood predictors in Tables 7 and 8, which was 22% in the human data 37% in the CDP++ Italian simulation data.

Figure 8

The impact of dominant-stress-neighborhood proportion in the human data and CDP++ Italian simulation



Note. The lines represent estimated intercepts and trends for the probability of dominant stress patterns assigned as a function of dominant-stress-neighborhood proportion in the human data (A) and CDP++ Italian simulation (B).

In general, with respect to the sublexical cues, it seems that the sublexical route's TLA network could pick up on those associations in principle, although it was not entirely successful at reproducing the pattern of results suggested by the human data. For example, the fact that the model was able to pick up an association between suffixes and stress although the model had no explicit rule system mediating those associations (unlike models such as Rastle and Coltheart's, 2000) or access to semantics suggests that that association was learned from the co-occurrence of suffixes and stress patterns. Indeed, in the CDP++ Italian simulation data as well as in the human data, the proportion of dominant stress patterns assigned varied as a function of the presence of dominant-stress suffixes (human: 81% with suffix present, 63% with suffix absent; CDP++: 83% with suffix present, 54% with suffix absent) and the presence of nondominant-stress suffixes (human: 55% with suffix present, 82% with suffix absent; CDP++: 44% with suffix present, 83% with suffix absent). Remember that similar proportions of dominant stress patterns were observed in the corpus analysis, albeit with a more pronounced difference for nondominant-stress suffixes (for dominant-stress suffixes, 80% with suffix present, 41% with suffix absent; for nondominant-stress suffixes, 19% with suffix present, 85% with suffix absent). Overall, this pattern of results seems to reflect a stronger tendency for both humans and CDP++ Italian to assign dominant and nondominant stress, respectively, when the corresponding suffixes are present. In other words, suffix effects also seem to reflect the learning of statistical probabilities related to stress.

However, these data do not help explain why the dominant-stress suffix effect was significant in the CDP++ Italian simulation data and not in the human data. A possible explanation might be that CDP++ Italian was more sensitive to stress neighborhood than humans, as noted above. Because suffixes would not be treated any different from any other ending by CDP++ Italian (because of the lack of a morphological/semantic system in the model), it would seem to follow that CDP++ Italian should be particularly sensitive to suffixes as well (although, if this idea represented the complete explanation, we would have expected an effect of nondominant-stress suffixes also to emerge in the simulation data).

Overall, a couple of results from the CDP++ Italian simulation data, while indicative of CDP++ Italian's capabilities, showed a bit of discrepancy with respect to the human data. First, the hierarchical regression analyses suggested that the specific association CDP++ Italian learned was that involving dominant-stress suffixes, which was present in the present corpus analysis but was not reflected in the human data. Similarly, with respect to the lexical cues, while it is clear from the many lexicalization errors that the lexical route had an impact in the simulation, that impact does not appear to involve a role for the source word's stress pattern and moderating factors.

General Discussion

Summary of results

A current mandate in reading research is that researchers should move beyond the examination of only monosyllabic stimuli in order to reach a more complete understanding of how words are read, as words in almost all languages are mainly polysyllabic (Mousikou et al., 2017). Doing so requires understanding the processes regulating stress assignment, especially in languages such as English or Italian in which the position of stress within a word is not readily predictable by rules, although cues to stress exist in those languages that readers can potentially use.

In the present project, we addressed this issue for Italian conducting, first, a corpus analysis (Study 1) which, confirming and extending the extant literature on cues to stress in Italian and other languages, revealed the existence of several sublexical units in the Italian vocabulary that readers might use as effective cues to stress assignment if they pick up on the vocabulary's statistical regularities (in addition to stress rules and knowledge of individual words' stress patterns). Next, we conducted a megastudy (Study 2) in which, unlike in standard factorial experiments, a large number of nonword stimuli was created involving variation in those stress cues as well as lexical cues (i.e., the resemblance of the nonword to its source word, the word used to create the nonword) which readers might also use. Finally, in a simulation using CDP++

Italian (Perry et al., 2014), a computational model of Italian polysyllabic reading, we evaluated the model's ability to reproduce the pattern of results suggested by the human data.

The results of the megastudy, largely in line with those of the corpus analysis, confirmed and extended extant knowledge on the cues that readers do use for stress assignment in Italian and other languages. For the sublexical cues, we found evidence for, first, stress dominance, a general preference for assigning the dominant stress pattern in Italian, i.e., penultimate stress (Colombo, 1992). This preference was observed not only for nonwords with a heavy penultimate syllable, a type of stimuli for which dominant stress is virtually always assigned in Italian, but also for nonwords with a light penultimate syllable for which stress is not as predictable. Second, for the latter type of nonwords, although dominant stress was generally preferred, this preference was modulated by several sublexical and lexical factors.

In particular, we confirmed a strong role for stress neighborhood, defined in terms of either the number of dominant-stress words sharing the nonword's ending (the dominant-stress-neighborhood count) and, even more strongly, the proportion of such words among the set of words sharing the nonword's ending (the dominant-stress-neighborhood proportion): Nonwords with a low/high dominant-stress-neighborhood proportion and/or count were less/more likely, respectively, to receive dominant stress, with proportion having a stronger effect (Burani & Arduino, 2004; Burani et al., 2014; Colombo, 1992; Sulpizio et al., 2013). As is apparent in Figure 4, the effect was more pronounced for nonwords with a low dominant stress neighborhood (in the left part of the graphs), as the bias for dominant stress led to a ceiling effect for nonwords with a dominant stress neighborhood (in the right part of the graphs).

Such a strong role for the endings does not seem to hold for English. Although some authors have proposed that endings are important (Arciuli & Cupples, 2006, 2007; Mousikou et al., 2017; Smith & Baker, 1976), more recently Ktori et al. (2018) argued that there is little evidence for the role of endings in English when other confounding factors are accounted for or eliminated. This difference for English may depend on the difference in complexity in syllable subdivision between English and other languages, which is crucial for the attribution of lexical stress. That is, syllable subdivision is quite straightforward in Italian (as well as being taught in primary school), while it is much more complex in English (for comparable evidence in Russian, see Jouravlev & Lupker, 2015b).

We also found a role for other sublexical cues to stress that our corpus analysis had revealed but that had not been examined previously in experiments for Italian (and that have received little attention for other languages as well). One such cue is the extent to which the specific syllable appearing in the penultimate position attracts stress, with nonwords having a penultimate syllable that, in the lexicon, is more often stressed than unstressed (i.e., nonwords with a higher penultimate-syllable stress proportion) being more likely to receive stress on that syllable. This result confirms previous speculations that readers may use distributional information specific to individual syllables in assigning stress, with typically stressed syllables being more likely to attract stress (Sulpizio et al., 2017). Note that this information is more specific than similar information that has been examined for other languages such as English (e.g., Guion et al., 2003; Kelly, 2004; Kelly et al., 1998; Mousikou et al., 2017; Treiman et al., 2020) for which the examination has typically concerned general characteristics of syllables such as onset and coda complexity (but see Jouravlev & Lupker, 2015b). The present results highlight the possibility that, in other languages as well, differences in the ability to attract stress may emerge between individual syllables matched on other relevant characteristics.

Affixes are other sublexical cues to stress revealed by our corpus analysis that have not been examined previously in experiments in Italian and that have received relatively little attention in other languages besides English, perhaps due to the difficulty of disentangling morphological and orthographic cues to stress, especially in morphologically rich languages. In contrast, for English, there has been an interest in the role of affixes on stress assignment since Rastle and Coltheart's (2000) computational model of English disyllabic reading, which includes an affix look-up procedure prior to stress being assigned (see also Kelly, 2004; Ktori et al., 2016, 2018; Treiman et al., 2020). Our megastudy results suggested a role for nondominant-stress suffixes (i.e., suffixes that, in Italian, are associated with nondominant stress) after controlling for the other predictors used in the model: Nonwords containing endings corresponding to those suffixes were less likely to receive dominant stress.

Do note that this suffix effect (as well as the effects of the other stress cues) was not extreme (e.g., a nonword containing a nondominant-stress suffix received dominant stress merely less frequently, not never or almost never). This type of result is consistent with the idea that what participants were using in assigning stress, particularly with the sublexical cues, was the learned associations between those cues and stress patterns existing in the language. Note however that this type of statistical-learning behaviour occurred alongside a process that could not be described in those terms, i.e., the one involved in the lexical cue use (as will be discussed in the next section).

Finally, it is worth noting that the analyses of the reading latencies, undertaken using a more exploratory approach, also suggested a role for (sublexical and lexical) cues that are typically associated with stress. After controlling for the effects of length (which showed a linear increase in latencies for stimuli up to 13 letters) and orthographic neighborhood, we found reduced latencies for nonwords with a heavy penultimate syllable, nonwords with a high number of dominant-stress neighbors, and nonwords with a dominant-stress source word, especially when the nonword resembled its dominant-stress source word the most. As these factors were also found to be important in the analyses of stress assignment, the implication seems to be that factors that make assignment of stress patterns more straightforward for certain stimuli (e.g., recognition that the penultimate syllable of a nonword is heavy, a type of syllable that virtually always bears stress) may also speed up reading of those stimuli. The reason would essentially be that, the easier the determination of a stress pattern, the earlier stress can be assigned and pronunciation can be programmed. Given the exploratory nature of our latency analyses, however, this idea would need to be examined more fully in future research undertaking a more confirmatory approach concerning the relationship between cues to stress and reading latencies.

Finally, the CDP++ Italian simulation revealed that CDP++ Italian captured some of the patterns suggested by the human data, i.e., the general bias for assigning the dominant stress pattern and the effect of stress-neighborhood proportion. However, it failed to capture several other patterns, including the effects of dominant-stress-neighborhood count, penultimate-syllable proportion, and all of the lexical cues. We discuss how these shortcomings might be addressed in the section “Inconsistencies between humans and CDP++ Italian” below, after having discussed the inconsistencies between, first, the present corpus and megastudy results and, second, between the present results and those produced by factorial studies in the relevant literature.

Inconsistencies between corpus and megastudy results

Although our megastudy results suggest that Italian readers assign stress using most of the sublexical cues revealed by our corpus analysis, there were some cues our participants did not seem to use. For some of those cues, the fact that they would not be used appears reasonable because the impact of those cues in the corpus analysis was actually quite small (i.e., the tendency for longer words to be less likely to have the dominant stress pattern) or unclear (e.g., the impact of the syllabic count measures) to begin with. For other cues, the fact that they were not used by our participants is a bit more surprising and worth noting. In particular, the fact that the effect of dominant-stress suffixes was not significant (while an effect of nondominant-stress suffixes did emerge, as noted above) seems to suggest that, for Italian readers, the presence of a dominant-stress suffix does not add much to the information that stress neighborhood already provides. Thus, for example, “-ato” is effective at cueing the dominant stress pattern not because it is a dominant-stress suffix, but because it is an ending associated with a stress neighborhood in which dominant stress prevails just as “-oro”, a non-suffix ending.

The reason for this insensitivity to dominant-stress suffixes may have to do with the fact that participants were already biased to assign the dominant stress pattern. For example, ambiguous suffixes (i.e., suffixes that were associated with both penultimate and antepenultimate stress, such as “-ano”) received a prevalence of dominant stress responses from our megastudy participants (i.e., 67%). For comparison, in our corpus, words with that type of suffix had dominant stress in a *minority* of cases (33%). In general, these discrepancies between corpus and megastudy results suggest that although readers can extract numerous statistical regularities relevant to stress from the vocabulary, they do not inevitably use all of them, likely because they prioritize those which are perceived as being more salient and/or reliable.

Further, the fact that effects of *lexical* cues emerged in our megastudy suggests that learning of statistical regularities is not sufficient to describe stress assignment performance in Italian and, potentially, other languages (note that, in our design, nonwords shared the ending with their source words, thus controlling for stress neighborhood, an important sublexical stress cue acquired through statical learning). Lexical factors are perhaps the most understudied factors in stress assignment, although some evidence from experiments in Greek suggests that nonwords tend to be assigned the same stress pattern as the words they resemble (Protopapas et al., 2007). Note also that dual-route computational models of reading do include a lexical route that at some point interacts with the sublexical route to determine stress (Perry et al., 2010, 2014; Rastle & Coltheart, 2000). With respect to Italian, it has been proposed that, in reading polysyllabic words, lexical phonology, and with it, the corresponding stress pattern, is automatically activated, and competes with or supports the sublexical information provided by the different types of cues (Colombo, 1992; Colombo & Sulpizio, 2015; Colombo & Tabossi, 1992; Sulpizio & Colombo, 2017; Colombo & Sulpizio, 2021).

Our megastudy results suggest that this activation may not be restricted to words but may occur with nonwords as well, as suggested by the emergence of an effect of the stress pattern of the word from which the nonword was derived. Interestingly but not unexpectedly, this effect was modulated by the similarity between the source word and the nonword as measured by the proportion of letters replaced in the former to create the latter: The higher the similarity (i.e., the lower the replaced-letter proportion), the more likely it was for the nonword to be stressed in the same way as its source word is. This pattern, however, was driven exclusively by nonwords with a nondominant-stress source word. For nonwords with a dominant-stress source word, the similarity with the source word had no influence on stress assignment, presumably because of a ceiling effect. However, in the pattern of latencies, explored in the Supplementary Materials, nonwords with a dominant-stress source word were read faster than nonwords with a nondominant-stress source word. This overall pattern suggests that dominant stress is computed quickly (Colombo & Sulpizio, 2021) and therefore the information from the source word and its relative similarity is not inevitably exploited.

A similar interaction pattern emerged when examining the frequency of the source word (i.e., nonwords with a more frequent, and hence more salient, source word were less likely to receive dominant stress when the source word was a nondominant stress source word) but the pattern was not significant in that case. The main point, in any case, is that stress assignment performance may not be merely the result of the use of (some) statistical regularities involving sublexical cues extracted from the vocabulary. It may also be a function of the resemblance of the nonword to a lexical representation, suggesting a role for similarity-based processes in stress assignment (Spinelli et al., 2016; see also the section "Inconsistencies between humans and CDP++ Italian" below).

Inconsistencies with factorial studies

With respect to the megastudy results as a whole, some additional issues do need to be addressed. The first concerns potential inconsistencies between the present results and those of previous studies. Some factorial experiments on Italian stress assignment focused on the contrast between stress dominance on one hand and other stress cues on the other hand, with some of them showing little or no preference/advantage for the dominant stress pattern. For example, when controlling for stress neighborhood, dominant-stress words are not read any faster than nondominant-stress words (Burani & Arduino, 2004; Burani et al., 2014; see also Colombo & Zevin, 2009), and dominant stress is not inevitably preferred when reading nonwords (Colombo et al., 2014; Sulpizio et al., 2013). In contrast, participants in our megastudy showed a general preference for assigning dominant stress, with cues that reinforced that pattern (e.g., a dominant-stress suffix or high similarity with a dominant-stress word) having little influence, whereas cues that suggested another pattern (e.g., a nondominant-stress suffix or high similarity to a nondominant-stress word) having a stronger influence.

This discrepancy between the results of the present megastudy and those of relevant factorial experiments in the literature may be explained by the fact that in those experiments, with one potential exception, there were no more stimuli bearing (in the case of words) or favoring (in the case of nonwords) dominant stress than stimuli bearing/favoring nondominant stress, unlike the normal situation in Italian in which dominant stress words outnumber nondominant stress words. The potential exception is Burani and Arduino's (2004) Experiments 1 and 2 in which the critical three- and four-syllable words were mixed with filler disyllabic words, words that always bear dominant stress unless the final syllable is marked with a diacritic. However, the facts that those filler words were disyllabic (as opposed to having three or four syllables as the critical words) and that many nonwords with indeterminate stress were also used as fillers make it unclear that a context favoring dominant stress overall was created in those experiments.

The implication of dominant stress not having been clearly favored in some factorial experiments is that those experiments may have produced list context effects (e.g., Lupker et al., 1997) that reduced the processing advantage/preference for dominant stress that would normally be observed in experiments, such as the present megastudy, in which the distribution of stress cues of Italian is more closely represented. Future studies should allow a more complete examination of this idea by, for example, manipulating the prevalence of stimuli bearing/favoring dominant stress within a trial block. The contrast between our results and those of factorial experiments, in any case, highlights the importance of the megastudy approach in examining processes in conditions with unbalanced probabilities of occurrence, probabilities that more closely reflect the situation in the normal experience.

Inconsistencies between humans and CDP++ Italian

As noted, the comparison between the human data and the CDP++ Italian simulation data revealed a few challenges for CDP++ Italian in terms of its stress assignment performance (for evidence along the same lines from the latency data, see the Supplementary Materials). Specifically, for the stress-neighborhood predictors, while the model produced a stress-neighborhood *proportion* effect, it did not produce a stress-neighborhood *count* effect, a failure that echoes that encountered by Perry et al. (2014) when attempting to simulate Burani and Arduino's (2004) Experiment 2 on the impact of the number of a word's stress friends. Further, while CDP++ Italian showed some sensitivity to morphological predictors, this sensitivity did not align with that suggested by humans because, while the latter showed, if anything, use of nondominant-stress suffixes on top of stress neighborhood information (although only when considering all of the nonwords and not just those compatible with CDP++ Italian), the model showed use of dominant-stress suffixes instead. Finally, CDP++ Italian showed no evidence that either the syllabic predictors or the lexical ones had an impact on the stress patterns it assigned.

None of these challenges seems to be insurmountable because, as noted in Study 3, the model should have the capability to use sublexical and lexical information to assign stress in similar ways as suggested by the human data. It is clear, however, that some changes would be necessary in order to produce that type of behavior. A minimal one would be to replace the corpus that Perry et al. (2014) used to train CDP++ Italian with the typical corpus used in Italian psycholinguistic studies, including the present one (i.e., CoLFIS and derived datasets; Bertinetto et al., 2005). It is possible that such a change would enable to model to capture the impact of the sublexical predictors it did not perform well with, such as stress-neighborhood count and the syllabic and morphological ones, whose associations with stress patterns might not have been as clear in the training corpus Perry et al. (2014) used to train CDP++ Italian.

Concerning the lexical predictors, a more substantial change would seem to be required. For example, the model parameters might be modified, first, in order for the main influence of the lexical route to be on the determination of the stress pattern rather than being the main determinant of the phonemes for the stimulus to be read (which is what seems to be the case with CDP++ Italian's default parameters, as all of the lexicalization errors it produced involved adding, substituting, or deleting phonemes from the nonword to be read; note that this type of errors were relatively rare in the human data). Second, this influence would have to be felt not only when the stimulus to be read is a perfect match to one of the stored lexical representations (i.e., when it is a word) but also when it is a close match (e.g., when it is a nonword similar to a word), an analogical process (Eddington, 2000; Glushko, 1979; Guion et al., 2003; Protopapas et al., 2007; Spinelli et al., 2016). Future research should consider exploring these possibilities in the light of the results of the present research.

Conclusion

In general, the present project can serve future research in several ways. For example, some of our corpus and megastudy results can be followed up in studies using the standard factorial approach in which the focus is on a smaller set of variables, with other relevant variables being carefully controlled when creating/selecting the stimuli. Further, the large set of stimuli that we used in the megastudy can be re-used, in whole or in part, in other experiments, including experiments targeting different populations such as developing readers or relevant patients such as dyslexic patients for whom there might be a different weighting of the cues used to assign stress (e.g., Arciuli et al., 2010; Colombo et al., 2014; Ktori et al., 2016; Rusconi et al., 2004). The behavioral dataset itself can also be explored in additional analyses aimed at examining the processes involved in human stress assignment. Finally, the human stress assignment performance examined in the present research can serve as a benchmark for the revision of computational models of polysyllabic reading in Italian (Pagliuca & Monaghan, 2010; Perry et al., 2014), potentially suggesting modifications to those models, and/or inspiring the creation of new models. For example, models such as the Italian version of the CDP++ (Perry et al., 2014), which has been developed for different languages using the same general architecture and learning principles across languages, may not be able to closely simulate the present results due to the fact that differences exist in the potential stress cues across languages. In contrast, a model in which stress assignment is viewed as a problem of Bayesian probabilistic inference and potential stress cues in the relevant language are used to estimate probabilities of stress patterns (Jouravlev & Lupker, 2015a) may fare better. It is hoped that future research will pursue the many avenues that the present megastudy opens up.

Notes

1. In the present manuscript, we refer to "rules" only when discussing *psychological* models that explicitly refer to the relevant processes as such, as opposed to *linguistic* models (e.g., generative ones) which may invoke "rules" to describe certain structural characteristics of a language. Note that a certain linguistic pattern might be called a "rule" in the linguistic sense but might not have a corresponding "rule" in the psychological sense. Indeed, the two computational models of Italian we are going to discuss (Pagliuca & Monaghan, 2010; Perry et al., 2014) lack an explicit rule system. As those models are based on orthographic and phonological layers, not connected by specific rules, their functioning reflects the process of learning the statistical properties of the language.
2. We used length in syllables rather than in letters as a predictor because syllables are theoretically more relevant than letters for the issue of stress assignment (as stress is a syllabic property).
3. Affixes are initial (i.e., prefixes) or final (i.e., suffixes) sequences within a word which contribute to the meaning of the word. As this definition implies, affixes are semantic in nature. However, semantics is not relevant in the present research because the stimuli we focus on in the megastudy (i.e., nonwords) are not meaningful ones, nor does the computational model we will use for the comparison with human data have access to semantics. Therefore, we will refer to affixes (prefixes and suffixes) simply as orthographic sequences which *can* contribute to the meaning of the stimulus they appear in.
4. Note that this definition of ending (i.e., the last 3 letters) differs from the traditional one discussed above (i.e., the graphemes going from the vowel of the penultimate syllable to the end of the stimulus; Colombo, 1992). The reason we used this definition is that, first, the graphemes going from the vowel of the penultimate syllable to the end of the stimulus typically *are* the stimulus's last 3 letters because VCV endings are the most common in Italian (Spinelli, Sulpizio et al., 2017). Second and most importantly, that

definition is the one Spinelli, Sulpizio et al. (2017) adopted in constructing Q2Stress, the database we used to extract stress-neighborhood information.

5. An anonymous reviewer of a previous version of the present manuscript wondered whether our finding that vowels in stressed syllables had a longer duration than vowels in unstressed syllables could be partially explained by the number of phonemes involved in the relevant syllables. In particular, assuming that participants would give approximately equal duration to each of the syllables in a nonword, there would be more time to articulate vowels in syllables with fewer phonemes (i.e., fewer consonants in the onset and/or coda in addition to the vowel nucleus) than in syllables with more phonemes. As a result, the vowel in the former syllables would a have longer duration and be more likely to be judged as being stressed even though the participant did not intend to place any emphasis on that vowel. In other words, our stress judgments might have been biased. To address this concern, we focused on the trials in which the nonword received stress on either the penultimate or the antepenultimate syllable (i.e., 1,796 trials out of 1800, the vast majority) and compared the number of phonemes involved in those syllables. Whether the stressed syllable had more, fewer, or the same number of phonemes as the unstressed syllable made hardly any difference in terms of the proportion of cases in which the stressed syllable had a longer duration than the unstressed one, proportions which were all in the 95–97% range. Thus, even if participants might have pronounced the syllables within a nonword at a regular pace and their doing so might have affected the relevant vowel durations in some cases, these results suggest that it is unlikely that that fact could have had a large impact on our stress judgments.
6. The reason that these cases (in particular the substitutions, additions, deletions, and transpositions) were considered errors is that since Italian has a transparent orthography, a nonword typically only has one acceptable representation at the segmental level (vs. several acceptable representations at the suprasegmental level, i.e., several acceptable stress patterns). (A potential exception is the pronunciation of mid vowels in stressed syllables, which can be either open, i.e., /ɛ/ and /ɔ/, or close, i.e., /e/

and /o/. However, this distinction does not seem to be perceived clearly by most Italian speakers who are not from either the Tuscany or Lazio regions in central Italy (note that the data for our megastudy were collected in Northern Italy). Further, even for speakers who do perceive that distinction, a mid vowel being pronounced as close instead of open (e.g., /'bene/ instead of /'bɛne/ 'good') or vice versa would not be considered an error. Therefore, we ignored that distinction for the purposes of the present research.) Any deviation from the acceptable representation does mean that an error must have occurred either in reading or in pronouncing the nonword (a type of error that, with unfamiliar stimuli as long as ours, is understandable). The situation is different in other languages such as English in which a nonword typically has several acceptable representations at the segmental level (e.g., Mousikou et al., 2017, recorded as many as 22 different pronunciations for a single nonword and even the computational models they examined rarely agreed on a single pronunciation).

7. Because the type of stimuli we used allowed an examination of secondary stress (an emphasis on a syllable that is weaker than that for the syllable bearing primary stress but stronger than that for unstressed syllables, such as for the syllable "in" in the word "independent"), an anonymous reviewer of a previous version of the present manuscript made the suggestion that we examine that type of stress, in addition to primary stress, our main focus. We have included an analysis of secondary stress in the Supplementary Materials. In a nutshell, we have concluded that the position of secondary stress, though somewhat hard to judge and not as acoustically distinct as the position of primary stress, is largely predictable based on the position of primary stress (Vogel & Scalise, 1982).
8. As noted, we did not predict an effect of prefix presence and did not include that predictor in the analysis. However, there was no indication in the data for an impact of prefix presence, as nonwords containing a prefix received dominant stress approximately as often (71.9%) as nonwords that did not contain a prefix (72.3%).

Acknowledgements

This research was partially supported by Grant DOR (Dotazione Ordinaria per Ricerca, Università degli Studi di Padova, year 2019) to Lucia Colombo, by Natural Sciences and Engineering Research Council of Canada Grant A6333 to Stephen J. Lupker, and by the Ontario Trillium Scholarship to Giacomo Spinelli. None of these funding sources had a role in study design, in the collection, analysis and interpretation of data, in the writing of the report, and in the decision to submit the article for publication.

References

Alfano, I., Schwab, S., Savy, R., & Llisterri, J. (2009). Cross-language speech perception: Lexical stress in Spanish with Italian and Francophone subjects. In *Atti del 5° Convegno Nazionale Associazione Italiana di Scienze della Voce* (pp. 455–474). Torriana, Italy: EDK Editore. Retrieved from https://www.aisv.it/PubblicazioniAISV/V_AISV/Articoli/Alfano_et_al.pdf

Andrews, S. (1986). Morphological influences on lexical access: Lexical or nonlexical effects?. *Journal of Memory and Language*, *25*, 726–740. https://doi.org/10.1016/0749-596X(86)90046-X

Arciuli, J., & Cupples, L. (2006). The processing of lexical stress during visual word recognition: Typicality effects and orthographic correlates. *Quarterly Journal of Experimental Psychology*, *59*, 920–948. https://doi.org/10.1080/02724980443000782

Arciuli, J., & Cupples, L. (2007). Would you rather 'embert a cudsert'or 'cudsert an embert'?. In A. C. Schalley and D. Khlentzos (Eds.), *Mental states: Language and cognitive structure volume 2: Language and cognitive structure* (pp. 213–237). Amsterdam, The Netherlands: John Benjamins.

Arciuli, J., Monaghan, P., & Seva, N. (2010). Learning to assign lexical stress during reading aloud: Corpus, behavioral, and computational investigations. *Journal of Memory and Language*, *63*, 180–196. https://doi.org/10.1016/j.jml.2010.03.005

Auguie, B. (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual word recognition volume 1: Models and methods, orthography and phonology* (pp. 90–115). New York, USA: Routledge.

Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers*, *34*, 424–434. https://doi.org/10.3758/BF03195471

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bertinetto, P.M. (1980). The perception of stress by Italian speakers. *Journal of Phonetics*, *8*, 385–395. https://doi.org/10.1016/S0095-4470(19)31495-0

Bertinetto P. M., Burani C., Laudanna A., Marconi L., Ratti D., Rolando C., & Thornton A. M. (2005). *Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS).* http://linguistica.sns.it/CoLFIS/CoLFIS_home.htm

Boersma, P. & Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.10. Retrieved from http://www.praat.org/

Bolker, B. (2022). *GLMM FAQ*. Retrieved from https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html

Boukadi, M., Zouaidi, C., & Wilson, M. A. (2016). Norms for name agreement, familiarity, subjective frequency, and imageability for 348 object names in Tunisian Arabic. *Behavior Research Methods*, *48*, 585–599. https://doi.org/10.3758/s13428-015-0602-3

Burani, C., & Arduino, L. S. (2004). Stress regularity or consistency? Reading aloud Italian polysyllables with different stress patterns. *Brain and Language*, *90*, 318–325. https://doi.org/10.1016/S0093-934X(03)00444-9

Burani, C., & Caramazza, A. (1987). Representation and processing of derived words. *Language and Cognitive Processes*, *2*, 217–227. https://doi.org/10.1080/01690968708406932

Burani, C., Marcolini, S., De Luca, M., & Zoccolotti, P. (2008). Morpheme-based reading aloud: Evidence from dyslexic and skilled Italian readers. *Cognition*, *108*, 243–262. https://doi.org/10.1016/j.cognition.2007.12.010

Burani, C., Paizi, D., & Sulpizio, S. (2014). Stress assignment in reading Italian: Friendship outweighs dominance. *Memory & Cognition*, *42*, 662–675. https://doi.org/10.3758/s13421-013-0379-5

Burani, C., Salmaso, D., & Caramazza, A. (1984). Morphological structure and lexical access. *Visible Language*, *18*, 342–352.

Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, USA: Lawrence Erlbaum. https://doi.org/10.4324/9781410606266

Colombo, L. (1992). Lexical stress effect and its interaction with frequency in word pronunciation. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 987–1004. https://doi.org/10.1037/0096-1523.18.4.987

Colombo, L., Deguchi, C., & Boureux, M. (2014). Stress priming and statistical learning in Italian nonword reading: Evidence from children. *Reading and Writing*, *27*, 923–943. https://doi.org/10.1007/s11145-013-9476-x

Colombo, L., Spinelli, G. & Lupker, S. J. (2020). The impact of consonant-vowel transpositions on masked priming effects in Italian and English. *Quarterly Journal of Experimental Psychology*, *73*, 183–198. https://doi.org/10.1177/1747021819867638

Colombo, L., & Sulpizio, S. (2015). When orthography is not enough: The effect of lexical stress in lexical decision. *Memory & Cognition*, *43*, 811–824. https://doi.org/10.3758/s13421-015-0506-6

Colombo, L., & Sulpizio, S. (2021). The role of orthographic cues to stress in Italian visual word recognition. *Quarterly Journal of Experimental Psychology*, *74*, 1631–1641. https://doi.org/10.1177/17470218211006062

Colombo, L., & Tabossi, P. (1992). Strategies and stress assignment: Evidence from a shallow orthography. In L. Katz and R. Frost (Eds.), *Orthography, Phonology, Morphology and Meaning* (pp. 319–340). Amsterdam: North-Holland.

Colombo, L., & Zevin, J. (2009). Stress priming in reading and the selective modulation of lexical and sub-lexical pathways. *PloS one*, *4*, e7219. https://doi.org/10.1371/journal.pone.0007219

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Domic (Ed.), *Attention and performance VI* (pp. 535–555). San Diego, USA: Academic Press. https://doi.org/10.4324/9781003309734-29

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Cortese, M. J. (2019). The megastudy paradigm: A new direction for behavioral research in cognitive science. In D. Spieler & E. Schumacher (Eds.), *New methods in cognitive psychology* (pp. 67–85). New York, USA: Routledge. https://doi.org/10.4324/9780429318405-4

Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology*, *66*, 946–972. https://doi.org/10.1080/17470218.2012.722660

Cutler, A. (2005). Lexical Stress. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 264–289). Oxford: Blackwell. https://doi.org/10.1002/9780470757024.ch11

Eddington, D. (2000). Spanish stress assignment within the analogical modeling of language. *Language*, *76*, 92–109. https://doi.org/10.1353/lan.2000.0022

Eriksson, A., Bertinetto, P. M., Heldner, M., Nodari, R., & Lenoci, G. (2016). The acoustics of lexical stress in Italian as a function of stress level and speaking style. In *Interspeech 2016* (pp. 1059–1063). https://doi.org/10.21437/interspeech.2016-348

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124. https://doi.org/10.3758/BF03195503

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression*. Thousand Oaks, USA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 674–691. https://doi.org/10.1037/0096-1523.5.4.674

Goslin, J., Galluzzi, C., & Romani, C. (2014). PhonItalia: a phonological lexicon for Italian. *Behavior Research Methods*, *46*, 872–886. https://doi.org/10.3758/s13428-013-0400-8

Guion, S. G., Clark, J. J., Harada, T., & Wayland, R. P. (2003). Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech*, *46*, 403–426. https://doi.org/10.1177/00238309030460040301

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*, 491–528. https://doi.org/10.1037/0033-295X.106.3.491

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662–720. https://doi.org/10.1037/0033-295X.111.3.662

Jouravlev, O., & Lupker, S. J. (2015a). Lexical stress assignment as a problem of probabilistic inference. *Psychonomic Bulletin & Review*, *22*, 1174–1192. https://doi.org/10.3758/s13423-015-0802-y

Jouravlev, O., & Lupker, S. J. (2015b). Predicting stress patterns in an unpredictable stress language: The use of non-lexical sources of evidence for stress assignment in Russian. *Journal of Cognitive Psychology*, *27*, 944–966. https://doi.org/10.1080/20445911.2015.1058267

Judica, A., Luca, M. D., Spinelli, D., & Zoccolotti, P. (2002). Training of developmental surface dyslexia improves reading performance and shortens eye fixation duration in reading. *Neuropsychological Rehabilitation*, *12*, 177–197. https://doi.org/10.1080/09602010244000002

Juphard, A., Carbonnel, S., & Valdois, S. (2004). Length effect in reading and lexical decision: Evidence from skilled readers and a developmental dyslexic participant. *Brain and Cognition*, *55*, 332–340. https://doi.org/10.1016/j.bandc.2004.02.035

Kelly, M. H. (2004). Word onset patterns and lexical stress in English. *Journal of Memory and Language*, *50*, 231–244. https://doi.org/10.1016/j.jml.2003.12.002

Kelly, M. H., Morris, J., & Verrekia, L. (1998). Orthographic cues to lexical stress: Effects on naming and lexical decision. *Memory & Cognition*, *26*, 822–832. https://doi.org/10.3758/BF03211401

Keuleers, E. (2013). vwr: Useful functions for visual word recognition research. R package version 0.3.0. https://CRAN.R-project.org/package=vwr

Ktori, M., Mousikou, P., & Rastle, K. (2018). Cues to stress assignment in reading aloud. *Journal of Experimental Psychology: General*, *147*, 36–61. https://doi.org/10.1037/xge0000380

Ktori, M., Tree, J. J., Mousikou, P., Coltheart, M., & Rastle, K. (2016). Prefixes repel stress in reading aloud: Evidence from surface dyslexia. *Cortex*, *74*, 191–205. https://doi.org/10.1016/j.cortex.2015.10.009

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lenth, R. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. https://CRAN.R-project.org/package=emmeans

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38. https://doi.org/10.1017/S0140525X99001776

Levy, R. (2014). Using R formulae to test for main effects in the presence of higher order interactions. *arXiv:1405.2094*. https://doi.org/10.48550/arXiv.1405.2094

Lüdecke, D. (2018). ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*, *3*, 772. doi:10.21105/joss.00772

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 570–590. https://doi.org/10.1037/0278-7393.23.3.570

Lupker, S. J., & Spinelli, G. (2023). An examination of models of reading multi-morphemic and pseudo multi-morphemic words using sandwich priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*, 1861–1880. https://doi.org/10.1037/xlm0001289

Lupker, S. J., Spinelli, G. & Davis, C. J. (2020a). Is zjudge a better prime for JUDGE than zudge is?: A new evaluation of current orthographic coding models. *Journal of Experimental Psychology: Human Perception and Performance*, *46*, 1252–1266. http://dx.doi.org/10.1037/xhp0000856

Lupker, S. J., Spinelli, G. & Davis, C. J. (2020b). Masked form priming as a function of letter position: An evaluation of current orthographic coding models. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *46*, 2349–2366. http://dx.doi.org/10.1037/xlm0000799

Mailhot, H., Wilson, M. A., Macoir, J., Deacon, S. H., & Sánchez-Gutiérrez, C. (2020). MorphoLex-FR: A derivational morphological database for 38,840 French words. *Behavior Research Methods*, *52*, 1008–1025. https://doi.org/10.3758/s13428-019-01297-z

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, *101*, 3–33. https://doi.org/10.1037/0033-295X.101.1.3

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*, 375–407. https://doi.org/10.1037/0033-295X.88.5.375

Mousikou, P., Sadat, J., Lucas, R., & Rastle, K. (2017). Moving beyond the monosyllable in models of skilled reading: Mega-study of disyllabic nonword reading. *Journal of Memory and Language*, *93*, 169–192. https://doi.org/10.1016/j.jml.2016.09.003

Napps, S. E. (1989). Morphemic relationships in the lexicon: Are they distinct from semantic and formal relationships?. *Memory & Cognition*, *17*, 729–739. https://doi.org/10.3758/BF03202634

Olivucci, F., Pasqualetto, F., Vayra, M., & Zmarich, C. (2016). Lo sviluppo dell’accento lessicale nel bambino in età prescolare: una prospettiva fonetico-acustica. In R. Savy & I. Alfano (Eds.), *La fonetica nell’apprendimento delle lingue* (pp. 209–228). Milan: Officinaventuno.

Pagliuca, G., & Monaghan, P. (2010). Discovering large grain sizes in a transparent orthography: Insights from a connectionist model of Italian word naming. *European Journal of Cognitive Psychology*, *22*, 813–835. https://doi.org/10.1080/09541440903172158

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*, 273–315. https://doi.org/10.1037/0033-295X.114.2.273

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, *61*, 106–151. https://doi.org/10.1016/j.cogpsych.2010.04.001

Perry, C., Ziegler, J. C., & Zorzi, M. (2014). CDP++. Italian: Modelling sublexical and supralexical inconsistency in a shallow orthography. *PloS one*, *9*, e94291. https://doi.org/10.1371/journal.pone.0094291

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56–115. https://doi.org/10.1037/0033-295X.103.1.56

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*, 859–862. https://doi.org/10.3758/BF03192979

Protopapas, A., Gerakaki, S., & Alexandri, S. (2007). Sources of information for stress assignment in reading Greek. *Applied Psycholinguistics*, *28*, 695–720. https://doi.org/10.1017/S0142716407070373

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rastle, K., & Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory and Language*, *42*, 342–364. https://doi.org/10.1006/jmla.1999.2687

Rusconi, M. L., Cappa, S. F., Scala, M., & Meneghello, F. (2004). A lexical stress effect in neglect dyslexia. *Neuropsychology*, *18*, 135–140. https://doi.org/10.1037/0894-4105.18.1.135

Sánchez-Gutiérrez, C. H., Mailhot, H., Deacon, S. H., & Wilson, M. A. (2018). MorphoLex: A derivational morphological database for 70,000 English words. *Behavior Research Methods*, *50*, 1568–1580. https://doi.org/10.3758/s13428-017-0981-8

Schiller, N. O., Fikkert, P., & Levelt, C. C. (2004). Stress priming in picture naming: An SOA study. *Brain and Language*, *90*, 231–240. https://doi.org/10.1016/S0093-934X(03)00436-X

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Ševa, N., Monaghan, P., & Arciuli, J. (2009). Stressing what is important: Orthographic cues and lexical stress assignment. *Journal of Neurolinguistics*, *22*, 237–249. https://doi.org/10.1016/j.jneuroling.2008.09.002

Smith, P. T., & Baker, R. G. (1976). The influence of English spelling patterns on pronunciation. *Journal of Verbal Learning and Verbal Behavior*, *15*, 267–285. https://doi.org/10.1016/0022-5371(76)90025-6

Spinelli, G., Sulpizio, S., & Burani, C. (2017). Q2Stress: A database for multiple cues to stress assignment in Italian. *Behavior Research Methods*, *49*, 2113–2126. https://doi.org/10.3758/s13428-016-0845-7

Spinelli, G., Sulpizio, S., Primativo, S., & Burani, C. (2016). Stress in context: Morpho-syntactic properties affect lexical stress assignment in reading aloud. *Frontiers in Psychology*, *7*, 942. https://doi.org/10.3389/fpsyg.2016.00942

Sulpizio, S., Arduino, L. S., Paizi, D., & Burani, C. (2013). Stress assignment in reading Italian polysyllabic pseudowords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 51–68. https://doi.org/10.1037/a0028472

Sulpizio, S., Burani, C., & Colombo, L. (2015). The process of stress assignment in reading aloud: Critical issues from studies on Italian. *Scientific Studies of Reading*, *19*, 5–20. https://doi.org/10.1080/10888438.2014.976340

Sulpizio, S., & Colombo, L. (2013). Lexical stress, frequency, and stress neighbourhood effects in the early stages of Italian reading development. *Quarterly Journal of Experimental Psychology*, *66*, 2073–2084. https://doi.org/10.1080/17470218.2013.785577

Sulpizio, S., & Colombo, L. (2017). Early markers of lexical stress in visual word recognition. *Memory & Cognition, 45*, 1398–1410. https://doi.org/10.3758/s13421-017-0735-y

Sulpizio, S., Spinelli, G., & Burani, C. (2017). STRESYL: An Italian Stress-in-Syllables database for reading research. *Written Language & Literacy*, *20*, 80–103. https://doi.org/10.1075/wll.20.1.05sul

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, *7*, 263–272. https://doi.org/10.3758/BF03197599

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*, 638–647. https://doi.org/10.1016/S0022-5371(75)80051-X

Talamo, L., Celata, C., & Bertinetto, P. M. (2016). DerIvaTario: An annotated lexicon of Italian derivatives. *Word Structure*, *9*, 72–102. https://doi.org/10.3366/word.2016.0087

Thornton, A. M., Iacobini, C., & Burani, C. (1997). *BDVDB: Una base di dati sul vocabolario di base della lingua italiana* [BDVDB: A database for the Italian basic dictionary]. Rome, Italy: Bulzoni.

Treiman, R., Kessler, B., & Hensley, K. (2023). Number and syllabification of following consonants influence use of long versus short vowels in English disyllables. *Journal of Memory and Language*, *129*, 104399. https://doi.org/10.1016/j.jml.2022.104399

Treiman, R., Rosales, N., Cusner, L., & Kessler, B. (2020). Cues to stress in English spelling. *Journal of Memory and Language*, *112*, 104089. https://doi.org/10.1016/j.jml.2020.104089

Vogel, I., & Scalise, S. (1982). Secondary stress in Italian. *Lingua*, *58*, 213–242. https://doi.org/10.1016/0024-3841(82)90034-1

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. https://ggplot2.tidyverse.org

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*, 502–529. https://doi.org/10.1016/j.jml.2009.02.001

Zhang, D. (2017). A coefficient of determination for generalized linear models. *The American Statistician*, *71*, 310–316. https://doi.org/10.1080/00031305.2016.1256839

Zhang, D. (2022a). Coefficients of Determination for Mixed-Effects Models. *Journal of Agricultural, Biological and Environmental Statistics*. https://doi.org/10.1007/s13253-022-00507-0

Zhang, D. (2022b). rsq: R-Squared and Related Measures. R package version 2.5. https://CRAN.R-project.org/package=rsq

Zoccolotti, P., De Luca, M., Di Pace, E., Judica, A., Orlandi, M., & Spinelli, D. (1999). Markers of developmental surface dyslexia in a language (Italian) with high grapheme–phoneme correspondence. *Applied Psycholinguistics*, *20*, 191–216. https://doi.org/10.1017/s0142716499002027

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1131–1161. https://doi.org/10.1037/0096-1523.24.4.1131