1-1-2014

Linearity and Word Internal Structure in the Visual Processing of Italian Complex Words

Franca Ferrari
QCC and Brooklyn College – CUNY, fferrari@qcc.cuny.edu

Natalie Kacinik
QCC and Brooklyn College – CUNY, NKacinik@brooklyn.cuny.edu

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/pwpl/vol20/iss1/11
For more information, please contact libraryrepository@pobox.upenn.edu.
Abstract
In this paper we address the theoretical debate about the representation and processing of derived words, presenting the results of a lexical decision task experiment aiming to investigate the visual recognition of Italian prefixed and suffixed words and nonwords. The study was specifically designed to test hypotheses stemming from the manipulation of two linguistic factors: (1) linearity (i.e., sequential order of morphemes), and (2) various degrees of complexity of the internal structure of particular types of Italian derived words.

Our findings show that suffixed and prefixed words and nonwords are not generally processed in the same way because they are not a homogenous set. Prefixes and suffixes instead engender different processing strategies, as has also been suggested by Cole et al. (1989) and Hay (2001). Moreover our data reveal that derived words are not necessarily processed more quickly or slowly than simple words (Bertram et al., 2000), but that their processing varies according to the nature of the affixes, the order in which they appear, and the information encoded in the affix.
Linearity and Word Internal Structure in the Visual Processing of Italian Complex Words

Franca Ferrari and Natalie Kacinik

1 Introduction

The access and processing of poly-morphemic words has been the subject of much debate. It is unclear whether such words are stored and accessed as (i) distinct morphemes (Taft and Forster 1975), (ii) whole words (Bybee 1985, 1985; Cole, Beauvillain & Segui, 1989), (iii) as both morphemic and whole forms accessed in parallel (Caramazza, Laudanna & Romani, 1988; Longtin & Meunier 2005, 2007; Rastle & Davis 2003; Schreuder & Baayen 1995; Taft 1994), along one or two pathways (Crepaldi et al., 2010, or Grainger & Ziegler, 2011, respectively); or (iv) whether morphemes are not represented discretely, but emerge through the interaction of semantic, phonological and orthographic codes (Gonnerman, Andersen, & Seidenberg, 2007; Plaut & Gonnerman, 2000; Rohde & Plaut, 2003).

A variety of linguistic factors have been studied to better understand the lexical access and processing of derived words, including (a) semantic and phonological transparency (Bertram et al., 2000); (b) change of the grammatical category (Marslen-Wilson, Bozic & Randall, 1997; Feldman, Barac-Cikola, & Kostic, 1998); (c) stem and surface frequency (Taft, 1979, Taft & Ardasinski, 2006); (d) language typology (Waksler, 2000); (e) configurational properties (Tsapkin, Jerema & Di Sciullo, 2004); (f) category information of the stem (Hudson & Buijs 1995); (g) word internal structure (Hudson & Buijs 1995); and (h) linearity (Cutler, Hawkins, & Gilligan, 1985; Hudson & Buijs 1995).

To address the theoretical debate about the representation and processing of derived words, we used a lexical decision task to investigate the visual recognition of Italian prefixed and suffixed words and nonwords. The study was specifically designed to test hypotheses stemming from the manipulation of two linguistic factors: (1) linearity (i.e., sequential order of morphemes, Hudson et al., 1995), and (2) various degrees of complexity of the internal structure of particular types of Italian derived words. According to the current literature, the present study appears to be one of the first to systematically investigate the interaction of suffixes and prefixes between and within three different kinds of word internal structure.

2 The internal structure of words and linearity

2.1 Factor 1: The internal structure of words

The internal structure of words depends on the types of combinations of affixes and stems that are available in a certain language. In Italian, for instance, there are two main types of affixes:

(A) Suffixes and prefixes that syntactically change the lexical category of the base they are affixed to and modify it semantically. We label this type as category changing affixes.

(1) fuma(re) base [V] + tore [N] > fumatore [N] (smoker)
(2) a [V] + largo base [Adj] (large) + are [V] > allargare [V] (enlarge)

(B) Suffixes and prefixes that do not syntactically change the lexical category of the base, but they modify the base only semantically. We label this type as non-category changing affixes.²

(3) piccolo (small) [Adj] + ino > piccolino [Adj] (smallish)
(4) scatola (box) [N] + ina > scatolina [N] (small box)

¹ The studies cited here are just representative due to pagination constraints. For a more detailed list of morphological processing studies readers are directed to Amenta and Crepaldi’s (2012) review article.
² Category changing affixes have also been labeled in the literature as ‘exocentric, external’ while non-category changing as ‘endocentric, internal’ (Di Sciullo, 1997).
(5) ri + comporre [V] > ricomporre [V] (recompose)

Given this distinction, it seems plausible to assume that category changing affixes generate internal structures of words that are configurationally and morphologically more complex than the structures generated by non-category changing affixes.

More specifically, category changing affixes determine the lexical category of the word and morphologically function as the head of a word, whereas non-category changing affixes are unable to function as heads, since the lexical category of the derived word is determined by its base. Moreover, the two types of affixes impose different lexical selection restrictions on their bases. Such restrictions require different structural configurations as illustrated in examples (6) and (7), respectively, for non-category changing and category changing prefixes.3

(6)

```
(ri) (fare) -> (rifare) (to redo)
```

(7)

```
(a(l)) (largo) -> (allargare) (to enlarge)
```

Looking at the representations in (6) and (7) each configuration is characterized by a shell with a different number of structural layers, such that Category changing affixes have an extra layer of structure in their shell. Regardless of whether the category change takes place in the lexicon, syntax or in a separate morphological component of the grammar, the internal structure of category changing affixed words is always characterized by an extra ‘bracket’, ‘layer’, or ‘node’ to accommodate the syntactic change. This suggests that the word internal structure of category changing affixed words is more complex (or has more layers) than the structure of non-category changing affixed words. Moreover, at a morphological level, category changing affixes are characterized by a three step word formation process -(i) affix-stem selection, (ii) changing the lexical class of the stem, and (iii) insertion of the affix morphology; whereas non category changing affixes only require a two-step process of (i) affix-stem selection and (ii) insertion of the affix morphology.

It is thus hypothesized that both the structural and morphological differences between these two types of affixed words affect their visual processing. At the other end of the complexity spectrum, the current study also included lexicalized Italian words formed with unproductive derivational affixes and/or bases that have lost their semantic transparency and acquired an idiomatic meaning.

(8) “ritrarre” (to retract), “pallino” (bullet)

---

3 Similar configurational differences distinguish non-category changing and category changing suffixes as well, but this is not represented in the present paper due to page limitations.
Although such lexicalized words appear to be structurally complex and may be subject to decomposition, at morphological and processing levels they are instead assumed to be represented, stored and accessed holistically like simple words as represented in (10) (Schreuder et al., 1994).

(10) V V Comp
ritrarre
(to retract)

### 2.2 Factor 2: Linearity

Linearity refers to the temporal linear order in which morphemes appear. This was investigated in the current study by comparing prefixed and suffixed words where the affix either precedes or follows the base, respectively. It has been claimed that ‘the temporal order of the detection of a particular type of violation indeed reflects processing stages’ (Friederici et al., 1996, p. 1222), and some research suggests that prefixed words may be privileged and undergo faster processing than suffixed words due to their position at the beginning instead of the end of a word, particularly when the prefix conveys information about the word’s lexical class (Friederici et al., 1996; Friederici & Meyer, 2004; Friederici & Weissenborn, 2007; Hagoort et al., 2003; Van den Brink & Hagoort, 2004). Specifically, these and other event related potential (ERP) studies have shown that the typical ELAN effect, obtained approximately 120-200 ms after the visual or auditory presentation of a word with a category violation, occurs earlier when “word category information was signaled by the prefix, but late with respect to word onset when word category information was provided by the suffix” (Friederici & Weissenborn, 2007, p. 52; see also Friederici, Gunter, Hahne, & Mauth, 2004; Hahne & Friederici, 2002).

In contrast, some behavioral studies of prefixed and suffixed words have obtained faster response latencies for suffixed words. These results have been explained in terms of ‘implicational and distributional’ language universals that show how there seems to be a preference for stems to precede affixes, with respect to both linguistic structure across languages (Hawkins & Cutler, 1988), and the processing preferences of individuals (Cutler et al. 1995). For instance, the cumulative effects of suffixed word frequencies have been found to result in faster reaction times (RT) for suffixed compare to prefixed word frequencies (Burani & Caramazza, 1987; Cole et al., 1989).

### 2.3 Nonwords

In the present study, the roles of and interaction between linearity and the internal structure of words were also tested with three types of Italian non-words created to mirror the 3 levels of internal structural complexity of the words. The 3 types of nonword stimuli: (1) regular nonwords, (2) pseudo words, and (3) mismatched words, were also selected to similarly reflect increasing levels of structural complexity, as outlined below. These stimuli were generally modeled upon the sorts of Italian nonwords used in previous research (e.g., Burani, Thornton, Iacobini, & Laudanna 1997; Burani, Dovetto, Spuntarelli & Thornton, 1999; Laudanna, Burani & Cemele, 1994).

(A) Regular non-words consisted of sequences of syllables created according to Italian phonological constraints but that do not correspond to real Italian words, e.g., *suttrico, *accova, *panido.

(B) Pseudo nonwords were formed with an Italian derivational prefixes or suffixes affixed to a nonword base, e.g., *in-gavire, *per-cotare and *prala-zione, *raccia-tore.

(C) Mismatched nonwords were created by taking Italian derivational prefixes or suffixes and affixing them to an Italian word whose lexical class does not match with the lexical class selected by the derivational morpheme, *in-grattare, *ri-pane and *calza[N]-mento[N], *casa[N]-zione[N].

### 2.4 Working Hypotheses

The current study was designed to test the following two hypotheses about the processing of de-
rived words and nonwords.

HYPOTHESIS 1. If the internal structural and morphological complexity of category changing affixed words is quantifiable in terms of the number of operations/steps/brackets/node involved in forming such words, then they should result in longer RTs and higher error rates relative to non-category changing affixed words and lexicalized words, which should be processed the fastest.

The rationale for this hypothesis is that lexicalized words should be the least structurally complex since the bound nature of the unproductive bases makes it unlikely that the bases would be stored separately as distinct words. It was therefore assumed that the lexicalized word stimuli would be stored and accessed holistically, similar to simple words (Schreuder et al., 1994). These words were thus expected to show the fastest processing times and greatest accuracy (Gurel, 1999).

With respect to the predictions for non-category changing and category changing affixed words, recall that the latter stimuli involve the additional operation or step required to change the grammatical category of the word (see 7-8 above). Category changing words were thus predicted to show the slowest reaction times and lowest accuracy, while responses to less complex non-category changing words were expected to fall somewhere in between the lexicalized and category changing stimuli.

Similar logic and predictions were expected for the nonword conditions, such that the fastest reaction times and fewest errors should be obtained for the least complex regular nonwords, followed by the pseudo and mismatched nonwords, respectively. The latter stimuli were deemed to be most complex since they consist of real, albeit mismatched, Italian affixes and stems thought to require a more thorough disambiguation process in order to recognize that the stimulus is not really a word. According to this rationale, it was assumed that pseudo nonwords would be easier to process and disambiguate because only the affix is a real component of Italian. This is similar to the reasoning made by Burani et al., (1987) in their study of Italian pseudo-affixed words.

HYPOTHESIS 2. The predictions made thus far have focused on the word internal complexity. The second hypothesis looks at the issue of linearity and the extent to which there may be differences between prefixed and suffixed words due to the position of the affix relative to the stem. Recall that Friederici and others have found that due to their position at the beginning of words, prefixes appear to privilege and result in faster word processing, particularly if the prefix provides information regarding lexical class (Friederici et al., 1996; Friederici & Weissenborn, 2007; Van den Brink & Hagoort, 2004). We therefore predicted that prefixed words would generally be responded to more quickly and accurately than suffixed words. However, if individuals are biased to prefer processing bases before affixes, then one should expect faster reaction times and better accuracies for suffixed words (Cole et al., 1989; Hawkins & Cutler, 1988; Hay, 2001). Another possibility is that the effects of linearity and word internal complexity will interact to affect prefixed and suffixed words in different ways. However, the most likely predictions with respect to the current literature are that slower RTs and lower accuracy should occur with increasing complexity – from lexicalized, to non-category changing, then category changing stimuli, and that generally faster RTs and higher accuracy should be obtained for prefixed compared to suffixed words, with a similar pattern of results expected for non-words.

3 Methods

3.1 Participants

The participants consisted of 120 right-handed native speakers of Italian, with an equal number of males and females. Seventy-five percent of the participants (n=90) were recruited and tested from various regions in Italy over two separate month-long data collection trips by the first author. The remaining 30 participants consisted of native Italian speakers who have been living in New York for a minimum of two years and a maximum of 15 years. Half of the participants (n = 60) were randomly assigned to run in either the prefix or suffix condition, which was treated as a between-subjects variable. Participants ranged between 20 to 50 years of age, with a mean age of 38 and they varied in their educational backgrounds, ranging from high school to post-graduate degrees such as Master and Doctorate degrees. Participation was done on a voluntary basis, such that the participants did not receive any compensation, monetary or otherwise, for their time.
3.2 Materials and Design

Twenty stimuli were generated for each of the 6 types of affixed words and nonwords along with 60 filler words, such that there were 180 items in the stimulus lists for both the prefix and suffix conditions. Across conditions, critical words were selected to be matched as closely as possible on the basis of length (in terms of both letters and syllables), language register, i.e., all words were extracted from the corpus CoLFIS (Corpus e Lessico di Frequenza dell’Italiano Scritto) based on written Italian excerpts from newspapers published between 1992-1994) and two measures of frequency. The first measure of word frequency was obtained from the SSLMIT, University of Bologna and Baroni, Marco, et al.’s “La Repubblica Corpus” consisting of 380 million token of Italian newspaper texts retrievable at the following web address: http://sslmit.unibo.it/repubblica. While the second measure of frequency was extracted by The corpus "PAISA" funded by ’Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR)', by means of the program Fondo per gli Investimenti della Ricerca di Base (FIRB) and retrievable at the following web address http://www.corpusitaliano.it/en/contents/paisa.html. The number of nouns, verbs, & adjectives was approximately equal across conditions. The regular nonword, pseudo- and mismatched nonword stimuli were created following the examples of similar types of items in the Italian and English experimental literature (Assink et al., 2000; Burani et al., 1995, 1999; Laudanna et al., 1994; Longtin & Meunier, 2005, 2007). All of the three types of nonwords were matched to the Italian word stimuli for length and the types of prefixes and suffixes used. Finally, the set of filler words consisted of simple Italian words and nonwords that also were also matched to the critical stimuli for length.

3.3 Procedure

After consenting to participate in the experiment, participants were randomly assigned to either the prefix or suffix condition. They were individually tested in a quiet room and seated 60 cm away from the screen of a 17 inch Samsung Q530 laptop computer equipped with DirectRT research software used to control stimulus presentation and record responses. Each stimulus item was displayed in the center of the screen in white Calibri 11-point font against a black background for 1000 ms followed by a blank 500ms inter-trial interval. Participants were instructed to make a lexical decision about whether each item appearing on the screen was a real Italian word or not by using their index and middle fingers to press the corresponding key as quickly and accurately as possible. For half of the participants, the left side arrow on the key board represented the “word” button and the down arrow key the “nonword” button, but these response keys were switched for the remaining 60 participants. The order of stimuli was randomized for each participant, and the experiment was run in four blocks of 45 trials with brief breaks in between each block.

4 Results

Overall, there was a significant main effect of condition for both RT, $F(1,118) = 1.49, p<.05$, and accuracy $F(1,118) = 14.41, p<.001$, such that suffix words and nonwords were generally responded to faster and more accurately than prefix stimuli.

4.1 Words

Both RT and accuracy analyses again showed a main effect of condition, $F(1,118) = 5.40, p<.05$, and $F(1,118) = 45.20, p<.0001$, respectively. The main effect of word type was not significant for RT, $F(1,236) = 1.09$, ns, but significant for accuracy, $F(1,236) = 30.31, p<.0001$. However, these main effects were qualified by a significant condition by word type interaction for both RT, $F(1,236) = 44.17, p<.0001$, and accuracy, $F(1,236) = 34.97, p<.0001$. 
Suffix results followed the expected trend of slower RTs with increasing complexity, although accuracy remained constant, whereas the Prefix results went in the opposite and unexpected direction with category changing and non-category changing words exhibiting the fastest and most accurate responses while lexicalized words were slowest and least accurate.

4.2 Non words

There was no main effect of condition in the RT data, $F(1,118) < 1$, and but it was significant for ACC, $F(1,118) = 4.75$, $p<.05$. The main effect of nonword type was also significant for both RT, $F(1,236) = 67.75$, $p<.0001$, and accuracy, $F(1,236) = 155.31$, $p<.0001$. The condition by nonword type interaction was significant for RT, $F(1,236) = 15.89$, $p<.0001$, but not for accuracy, $F(1,236) < 1$. 
Since responses to prefixed nonwords were generally slower and less accurate than suffixed nonwords, these results fail to support the linearity hypothesis. However, the nonword results provide partial support for the complexity hypothesis. Regular nonwords were processed faster and with fewer errors than the more complex nonword types, but contrary to predictions pseudo-words in both conditions were generally slower and less accurate than the mismatched words.

5 Discussion

The goal of this study was to examine the interaction between linearity (prefix vs. suffix) and three levels of internal structural complexity across different types of derived words and nonwords. Lexicalized words and regular nonwords were expected to be processed faster and with fewer errors based on the assumption of being the least structurally complex. It was further hypothesized that non category changing words and pseudo nonwords would be processed faster and more accurately than category changing words and mismatched nonwords, respectively, since the former stimuli remain in the same semantic category as their original stem and are less structurally complex. Finally, with respect to linearity, prefixed words were generally predicted to be processed faster and more accurately than suffixed words.

Contrary to the linearity hypothesis, however, the results showed that suffixed words were generally responded to faster and more accurately than prefixed items. This suggests a parsing preference for base-affix rather than affix-base sequences, a finding that has also been supported.
by other cross-linguistic studies of word recognition and production (e.g., Cole et al., 1989; Cutler et al. 1995; Hawkins & Cutler, 1988; and others). It also argues against the claim that prefixes undergo lexical-semantic interpretation prior to suffixes (Friederici et al., 1996, 2004; Friederici & Meyer 2004; Friederici & Weissenborn, 2007; Hagoort et al., 2003; Van den Brink & Hagoort, 2004). The nonword results also failed to support the linearity hypothesis since all types of prefixed nonwords were either less accurate than suffixed nonwords, or showed no difference in RT.

Linearity was additionally found to interact with internal structural complexity to produce an opposite pattern of results for prefixed and suffixed words, such that only the suffixed word effects went in the expected direction, with lexicalized suffixed words faster than non-category changing, followed by category changing words which were the slowest. The findings for prefixed words, on the other hand, went in the opposite direction, with the most structurally complex category changing prefixes likely privilege syntactic-category information such that it is processed first. This suggests that at a processing level individuals distinguish between syllabic incipits such as ri-, in-, il-, im- that are category changing from those that are simply unmarked as it is the case of lexicalized words. Based on Laudanna et al. (1994) and Laudanna and Burani (1995), speakers may favor category changing prefixes over the unmarked types because they are more likely to serve as processing units, and hence more salient.

The structural complexity hypothesis was also partially supported by the nonword results since regular nonwords were found to be faster and more accurate than both pseudo and mismatched stimuli. Contrary to our predictions, however, it was the pseudo-affixed rather than mismatched nonwords that were the least accurate in both conditions. Although the RT difference between pseudo and mismatched words was not significant for prefixed stimuli, suffixed pseudo words elicited significantly longer RTs than mismatched words. These results therefore fail to confirm those of previous studies with English prefixes that obtained longer rejection time for mismatched (dejuvenate) than pseudo nonwords (depertoire), and with Italian suffixed nonwords that showed no RT difference between mismatched and regular nonwords (e.g., Taft & Forster, 1975, & Burani et al., 2002, respectively).

We account for this finding of longer RTs for pseudo than mismatched suffixed nonwords by proposing that the parser can rule out grammatical errors arising from syntactic features mismatches more quickly than semantic errors. Our data also suggest that a real Italian affix attached to a non-word stem engenders a more exhaustive search to make sure the stem and the nonword stimulus as a whole isn’t really a word. This search results in a greater variability of responses in terms of longer RTs and lower accuracy, whereas mismatched items can be quickly rejected as not being a word as soon as the mismatch is detected with respect to either the affix or the base.

The asymmetric behavior between suffixed and prefixed words and nonwords indicates that affixed words are not generally processed in the same way, but that prefixes and suffixes instead engender different processing strategies, as has also been suggested by Cole et al. (1989) and Hay (2001). When processing derived words in general, it seems like there is a widespread preference for processing bases before affixes which results in faster recognition of suffixed than prefixed words. However, the response times and accuracies for suffixed words in particular, are affected by the complexity of the suffix, such that the parser processes lexicalized words the fastest since they are essentially 'bases', but has the greatest difficulty with the most complex category changing suffixes. On the other hand, while processing prefixed words the parser appears able to privilege the processing of syntactic information of the prefix prior to the base. Indeed, it seem like the more information the prefix can provide about the syntactic and semantic features of the word, the greater the efficiency and accuracy with which the word is processed. This also potentially explains why lexicalized words, whose incipits have been voided of both semantic and syntactic information, showed the longest RTs and highest error rates.

Indeed, the biggest differences between suffix and prefix conditions were obtained for the lexicalized words. This suggests that lexicalized words are not a homogeneous set, and provides further confirmation of the above conclusion that prefixed and suffixed words are processed differently. These findings thus indicate that derived words are not necessarily processed more quickly or slowly than simple words (Bertram et al., 2000), but that their processing varies according to
the nature of the affixes, the order in which they appear, and the information encoded in the affix.

6 Conclusion

In conclusion, the lack of homogeneity between suffixes and prefixes posits a challenge to traditional morphological classifications of affixes. Our study shows that the difference between prefixes and suffixes is not simply a linear difference with respect to their position before or after a base but that prefixes and suffixes of the same linguistic type may react differently to linguistic, psycholinguistic, and structural factors during word processing. Since suffixes and prefixes are not homogeneous, but instead appear to be stored, accessed, and processed differently, our findings demonstrate the inadequacy of strictly dichotomous claims that all derived words must be stored in either decomposed or whole word forms, (Cole et al., 1989; Taft & Forster, 1975), and accessed via one or two distinct routes (Caramazza et al., 1988).

The distinct nature of prefixed and suffixed words makes it difficult for strict left to right linear models to account for their processing differences, but connectionist constraint satisfaction theories of lexical processing (Plaut & Gomnerman, 2000; Seidenberg & MacDonald 1999) may provide a solution. Such theories operate under the assumption that all words are processed by the same network and that “morphological structures emerge from the convergence of other codes” (Mirkovic et al., 2011: 58), more precisely from the interaction between overlapping semantic, phonological, and orthographic codes. If words do indeed result from various overlapping levels of representation, then there may be no need “for a discrete, isolable level of morphological representation”. Such an account has the flexibility to potentially account for the variable pattern of results obtained in this study, although future research will have to precisely determine which dimensions and constraints are particularly critical for processing complex words in Italian.

References


Mirković, Jelena, and Maryellen MacDonald. C. 2013. When singular, and plural are both grammatical: Semantic, and morphophonological effects in agreement. *Journal of Memory, and Language*.


Department of Speech & Department of Psychology
QCC and Brooklyn College – CUNY
New York, NY
fferrari@qcc.cuny.edu, nkacinik@brooklyn.cuny.edu