

The Representation of Prefixed Forms in the Italian Lexicon: Evidence from the Distribution of Intervocalic [s] and [z] in Northern Italian*

Marco Baroni, UCLA

Abstract

In Northern Italian, the phones [s] and [z] are in complementary distribution and only [z] can occur in intervocalic position (for example, ca[z]la ‘house’). Although intervocalic /S/ voicing is a productive process, it does not apply to stem-initial /S/, which is realized as [s] (for example, a+[s]limmetrico ‘asymmetrical’). This morphological blockage of intervocalic /S/ voicing is also synchronically active.

Thus, the distribution of [s] and [z] provides evidence on how potentially prefixed forms are represented by Italian speakers. For example, Northern Italian speakers produce the (etymologically prefixed) word pre+/S/unto ‘presumed’ with [z], which indicates that they no longer represent it as prefixed. A survey of how 12 Northern Italian speakers produced words with (potentially) stem-initial /S/ was conducted. The survey results show that speakers represent some lexicalized forms as morphologically complex and that they oscillate between treating some forms as prefixed and non-prefixed. The likelihood that a form is treated as prefixed is significantly affected by the semantic transparency of its component parts and by the length of its prefix.

1. Introduction

An important issue in the study of morphology is to determine which forms are treated as morphologically complex by the speakers of a language. It is very likely that for most contemporary English speakers to re-fry is a complex form, and real is a monomorphemic form, even if it begins with the string re. However, in many cases the issue of whether a word is represented as morphologically complex or not is far from trivial.

Consider for example the word to resist. In this form, re- does not have its productive iterative meaning; the potential stem -sist is not an independent word of English, and it is hard to associate a meaning to it. On the other hand, -sist occurs in a number of other potentially prefixed forms (assist, consist, desist, insist, persist, subsist). Some of these forms are also (vaguely) semantically related to to resist. Thus, we cannot decide a priori whether to resist is a simple or complex form for English speakers.

The issue of how potentially morphologically complex forms are represented in the lexicon has been extensively investigated by psycholinguists using various experimental paradigms (see, among others, Henderson 1985 and the papers collected in Feldman (ed.) 1995 and Sandra & Taft (eds.) 1994).

In this study, I will present evidence on how potentially complex words are represented in the Italian production lexicon, and on which factors affect the speakers' decision to treat words as complex or simple. The evidence presented here comes from the analysis of a systematic morphophonological phenomenon of (Standard) Northern Italian.¹

In Northern Italian the alveolar fricatives [s] and [z] are in complementary distribution. In particular, only the allophone [z] can occur in intervocalic position (1.a); only the allophone [s] can occur word-initially before vowels (1.b).

(1) ²	a.	['kaza] 'home'	b.	['santo]	'saint'
		['vizo] 'face'		['sot:o]	'under'
		[ri'zata]	'laughter'	[so'lare]	'solar'
		*['kasa]		*['zanto]	

I refer to the fact that only [z] can occur in intervocalic position with the descriptive label of “intervocalic voicing”, and I adopt the symbol /S/ to refer to the alveolar fricative phoneme, not specified for [±voice]. As I will show in section 2, intervocalic /S/ voicing is a synchronically productive rule/constraint of Northern Italian.

Intervocalic /S/ voicing is sensitive to the morphological structure of words. As the examples in (2) show, if /S/ is stem-initial, it does not undergo intervocalic voicing:

- (2) [a+si'm:ɛtriko] ‘asymmetrical’
 [a+so'ʃale] ‘anti-social’

In these words /S/ occurs in intervocalic position but, since the vowel preceding the alveolar fricative belongs to a prefix, the fricative does not undergo intervocalic voicing. Thus, from the point of view of the distribution of [s] and [z], the stem-initial context behaves like the word-initial context (1.b). I will refer to this pattern as Intervocalic Voicing Blocking (IVB). I assume the following, purely descriptive definition of IVB:

- (3) Intervocalic Voicing Blocking
 Stem-initial /S/ is not subject to intervocalic voicing.

In section 2 I will show that IVB, like intervocalic voicing, is a productive pattern of Northern Italian.

Interestingly, intervocalic voicing is not blocked in morphological contexts different from that illustrated by (2). Thus, even if the only allophone allowed in word-final position is [s] (4.a), prefix-final (4.b) and stem-final (4.c) /S/s undergo intervocalic voicing:

- (4) a. ['bus] ‘bus’
 ['lapis] ‘lapis’
 *['buz]
- b. [diz+abitu'are] ‘to get somebody out of a habit’
 [ʃiz+al'pino] ‘that is located on this side of the Alps’

- c. [ˈspoz+a] ‘spouse + fem. sg. suffix’
 [naz+ˈino] ‘little nose (nose + diminutive suffix)’

It is also interesting to observe that IVB applies independently of the morphological structure of a word, as long as the word displays a stem-initial /S/. Consider for example the derivations in (5):

- (5) a. [senile]_A ‘senile’
 [pre [senile]_A]_A ‘pre-senile’
 [[pre [senil]_A]_A izzare]_V ‘to pre-senile-ize (to make something/somebody pre-senile)’
- b. [sociale]_A ‘social’
 [[social]_A izzare]_V ‘to socialize’
 [ri [[social]_A izzare]_V] ‘to re-socialize’

The words presenilizzare and risocializzare have different morphological constituency. In presenilizzare, the prefix pre- is attached to the adjectival stem senile, and then the verbalizing suffix -izzare is attached to the prefix-stem complex. In risocializzare, on the other hand, the prefix ri- is attached after the suffix -izzare, in the last stage of the word formation process. Nevertheless, since both words display a stem-initial /S/, they are both subject to IVB: they are produced as pre[s]enilizzare and ri[s]ocializzare, respectively.

The focus of the previous literature on IVB (Nespor & Vogel 1986; Vogel 1994; Peperkamp 1995; Kenstowicz 1995; Loporcaro 1996) has been on the problem of deriving the asymmetry between /S/ in stem-initial position, where IVB applies, and /S/ in other contexts, where intervocalic voicing occurs across morphemes. Different versions of intervocalic /S/ voicing and IVB have been proposed in order to account for this asymmetry; however, the claim made here that the words in (2) display an intervocalic [s] because they are morphologically complex is not controversial.³ The data presented in this study do not bear on the issue of how to account for the asymmetry between stem-initial /S/ and /S/ in other contexts.

Since intervocalic voicing per se is productive and exceptionless, intervocalic [s] only occurs in morphologically complex words as a product of IVB. Thus, the occurrence of intervocalic [s] can be used as a diagnostic of morphological complexity: if a potentially prefixed word with /S/ in (potentially) stem-initial position is realized with [s], this means that contemporary speakers treat the word as morphologically complex; if /S/ is realized as [z], then the morphological structure of the word is synchronically opaque. Consider for example the forms in (6):

- (6) [pre'zunto] 'presumed'
 [rezis'tente] 'resistant'

As the glosses suggest, these could be considered prefixed words in which the alveolar fricative is stem-initial. However, in these cases intervocalic voicing is not blocked. Intuitively, the reason for this is that the forms in (6) are not morphologically complex from a synchronic point of view. In contemporary Italian, their morphological structure is opaque, and speakers treat them as monomorphemic forms.⁴

In this research, I use the distribution of IVB as a tool to explore the way in which potentially complex forms with the relevant structure (V(+))SV are represented by contemporary Northern Italian speakers in their production lexicon, and the factors which induce speakers to treat forms as morphologically complex (at least for the purpose of IVB).

The results of this investigation suggest that speakers are not only aware of the morphological structure of forms respecting the productive word-formation patterns of Italian, but they also represent some lexicalized words in a morphologically complex format. Furthermore, in some cases speakers oscillate between a morphologically complex and a morphologically simple representation of the same words. The main factors which determine the likelihood that a form beginning with a string identical to a prefix is treated as

complex are the semantic transparency of its components and the phonological length of the potential prefix.

The remainder of this paper is organized in the following way: In section 2, I show that both intervocalic /S/ voicing and intervocalic voicing blocking are productive patterns of contemporary Northern Italian; in section 3, I describe the survey that I conducted in order to gather data on the distribution of [s] and [z] in potential stem-initial contexts. The results of the survey are analyzed in section 4.

2. The synchronic productivity of intervocalic /S/ voicing and IVB

The rules/constraints governing the distribution of the Northern Italian allophones [z] and [s] outside the IVB context are exceptionless and they are actively applied in the production of nonce forms and loanwords. This is true for [z] and [s] in all phonotactic contexts; however, here we are only interested in intervocalic [z] (7.a) and word-initial [s] (7.b):

- (7) a. ['kaza] 'home' b. ['santo] 'saint'
 ['vizo] 'face' ['sot:o] 'under'
 * [...VsV...] * [zV...]

While conducting a separate study on Northern Italian [s] and [z] (Baroni 1996), I asked 58 speakers to read a list of nonsense words including fillers and forms with <s>, which is the grapheme corresponding to both [s] and [z] in Italian orthography. All the speakers systematically produced /S/ as [s] in nonsense forms with word-initial <s> and as [z] in nonsense words with intervocalic <s>. For example:

- | | | |
|-----|---------------------------|-----------------------|
| (8) | <u>Orthographic form:</u> | <u>Pronunciation:</u> |
| | samo | ['samo] |
| | satto | ['sat:o] |
| | pasa | ['paza] |
| | laso | ['lazo] |

Analogously, loanwords are adapted to the Northern Italian distribution of [s] and [z]:

- | | | |
|-----|------------|---|
| (9) | [ˈsarland] | (from German ˈ[z]aarland)
‘German district’ |
| | [ˈsaker] | (from German ˈ[z]acher)
‘type of cake’ |
| | [ˈme:za] | (from Spanish or English ˈme[s]a)
‘mountain with flat top’ |
| | [elˈpa:zo] | (from Spanish or English ˈEl Pa[s]o)
‘name of town’ |
| | [aˈza:ki] | (from Japanese ˈA[s]ahi)
‘brand of beer’ |
| | [z̥ zaka] | (from Japanese ˈO[s]aka)
‘name of town’ |

Notice also that in other varieties of Italian [s] can occur in intervocalic position (Canepari 1992). Consequently, Northern Italian speakers could find a rich source of potential exceptions to intervocalic voicing (i.e. words with intervocalic [s]) in the speech of other Italians.

It seems safe to conclude that the distribution of [s] and [z] in Northern Italian is governed by synchronically active phonotactic constraints, and it is not simply the lexicalized product of diachronic trends. Similarly, IVB is not restricted to a list of lexical forms, but it systematically applies to new prefixed forms, as the following test shows.

10 Northern Italian speakers were asked to create a series of new denominal verbs from proper nouns by adding the suffix -izzare ‘-ize’ (this is a productive word formation process of contemporary Italian), and then to add the prefixes ri- ‘re-’ and dis- ‘de-’ to the resulting forms (both prefixes can be productively attached to verbs). The speakers had to repeat this procedure for a list of 20 proper nouns, including 3 items beginning with /S/ (10.a) and 3 items beginning with a vowel (10.b):

- (10) a. Sandro b. Aldo
 Serena Elisa
 Sergio Ezio

All the speakers produced intervocalic [s] when forming new denominal verbs with ri- and names beginning with /S/ (11.a), and intervocalic [z] when forming new denominal verbs with dis- and names beginning with a vowel (11.b):

- (11) a. ri+[s]andr+izzare 'to re-Sandro-ize'
 ri+[s]eren+izzare 'to re-Serena-ize'
 ri+[s]erg+izzare 'to re-Sergio-ize'
- b. di[z]+ald+izzare 'to de-Aldo-ize'
 di[z]+elis+izzare 'to de-Elisa-ize'
 di[z]+ez+izzare 'to de-Ezio-ize'

Intervocalic voicing in the forms with dis- shows that the blockage of voicing in the forms with ri- is not due to some kind of task-specific factor. Thus, the occurrence of [s] in the forms in (11.a) must be due to the productive application of IVB in a morphologically complex context.

Since intervocalic /S/ voicing is otherwise a productive Northern Italian pattern, but intervocalic /S/ is realized as [s] in nonce formations in which it is stem-initial, we can safely conclude that the morphology-dependent IVB rule is part of the linguistic knowledge of Northern Italian speakers, and it is not simply a linguists' generalization about a list of lexicalized forms with exceptional intervocalic [s].

3. IVB as a diagnostic of morphological complexity: the IVB survey

In the previous section we saw that only stem-initial /S/s are not subject to the otherwise productive and systematic process of intervocalic /S/ voicing. Thus, we can use IVB as a test of where Italian speakers posit prefix-stem structures (beyond the trivial case of nonce formations), and where they treat etymologically prefixed forms as monomorphemic.

For example, the words in (12) are always produced with intervocalic [s], which indicates that they are treated by the speakers as prefixed forms in which /S/ is stem-initial:

- | | | |
|------|--------------------|--|
| (12) | [pre+sele't:sjone] | 'preliminary selection' |
| | [a+si'm:ɛtriko] | 'asymmetrical' |
| | [a+so'tʃale] | 'anti-social' |
| | [pre+se'nile] | 'characteristic of the age before old age' |
| | | Cf. [se'nile] 'pertaining to old age' |

On the contrary, the three words in (12) are always produced with intervocalic [z], since their morphological structure is opaque to contemporary speakers:

- | | | |
|------|-------------|---------------|
| (13) | [de'zɛrto] | 'desert' |
| | [pre'zaɟo] | 'premonition' |
| | [pre'zunto] | 'presumed' |

In order to gather data on the morphological structure of Italian words using the IVB test, I organized a survey in which fourteen speakers had to read 108 forms with a (potentially) stem-initial /S/ aloud five times. The reason for collecting multiple repetitions of each word is that for some words, as the survey results confirm, speakers oscillate between [s] and [z] realizations. As a form of control on the validity of the data collected in this way, an acceptability judgment survey was also conducted (see discussion in 3.2).

In the psycholinguistic literature on morphological processing and lexical representation, it is common to distinguish between the modality specific units used in speech comprehension, the modality independent lexical storage units containing semantic,

morphological and syntactic information (“lexical entries”), and the units used in speech production (for distinctions along similar lines see, among others, Marslen-Wilson, Tyler, Waksler & Older 1994, Burani 1992, Caramazza, Laudanna & Romani 1988, Roelofs & Baayen 1996). Clearly, the IVB test provides evidence on the latter kind of units, i.e. on what I will label as “the phonological output lexicon” (following Burani 1992) or “production lexicon”.⁵

3.1 Survey design and data collection

A list of all the Italian prefixes ending in a vowel was prepared using the lists in Dardano (1978) as the basic reference. All the words in Zingarelli (1987) in which one of these prefixes was followed by a (potential) stem beginning with /S/ were collected. Since this list was too long, it was trimmed on the basis of the following criteria: only few words for each cluster of derivationally related words were kept; only a limited number of technical words and archaisms was kept; only a limited number of words that would be probably considered nonce formations by most speakers was kept. In this way, the list was reduced to 108 words. Ten control items were added: five words in which /S/ occurs intervocalically within the same morpheme; five clitic + host word phrases in which /S/ occurs at the beginning of the host word. The items in each control set are matched with some target word, so that in the corpus there are five triplets formed by a control with VSV inside a morpheme, a control with V##SV between a clitic and a word, and a target word with V+SV at the prefix-stem boundary. Within each triplet, each item has the same number of syllables, stress falls on the same syllable, and /S/ occurs in the same location and is surrounded by the same vowels.

Each word (and clitic-word sequence) was embedded in a carrier sentence. The carrier sentences were 25 to 35 characters long, and they did not contain /S/s other than those in the target/control words.

Words were embedded in carrier sentences for several reasons. First, in this way the task became more natural and closer to a real speech situation.

Second, the carrier sentences disambiguate the meaning of potentially ambiguous words. For example, risaltare means ‘to stand out’, but it could also be a nonce formation meaning ‘to jump again’ (from ri- and saltare ‘to jump’). However, once the word is embedded in the sentence Il giallo è un colore che risalta ovunque ‘Yellow is a color that stands out everywhere’, only the first meaning is plausible.

Third, carrier sentences had the function to create a time lag between the production of the target words, and to distract the speakers’ attention from them. This was especially important because no fillers were added.

The carrier sentences were designed so that the target words would occur in a natural context, but avoiding contexts in which their morphological structure would have become explicit. For example, there were sentences like: ‘Look at how asymmetrical that triangle is!’, but not sentences like: ‘If a triangle is not symmetrical, then it is asymmetrical’.

Most Italian words, and all the words in the survey, must carry an inflectional suffix. For each form, I selected what I felt to be the most common inflectional ending. This means that in most cases I selected the singular form of nouns, the masculine singular form of adjectives, and infinitival, present indicative or participial forms of verbs. However, in some cases I felt that other forms were less marked. For example, the adjective asintotico ‘asymptotic’ is typically used in the phrase curva asintotica ‘asymptotic curve’, where it appears in its feminine form, so this is the form in which the adjective was presented in the survey.

Each subject had to read the whole list of sentences five times. Each sentence appeared on a computer screen for 1500 milliseconds; the sentences were separated by 750 millisecond intervals. In order to minimize ordering effects, each time the list was presented the sentences appeared in a different random order (the order varied between presentations to the same speaker and across speakers).

Fourteen native speakers of Northern Italian took part in the survey. After the reading-and-recording session, the speakers read a list of nonsense words containing forms with intervocalic and word-initial /S/ and fillers. All of them systematically produced intervocalic /S/ as [z] and word-initial /S/ as [s]. After this test, the speakers were interviewed on their dialectal background and, finally, they were asked to fill out a form in which they had to rate the “semantic transparency” of each complex word presented in the test (see section 4.3).

The recordings of two speakers were discarded since it emerged from the post-recording interviews that they could not be considered standard speakers of Northern Italian. I listened to the recordings of the twelve remaining speakers, marking for each repetition of each word whether /S/ was produced as [s] or [z]. This impressionistic data analysis procedure is justified by the acoustic and articulatory data presented in Baroni (1998), where I showed that, at least for the speaker analyzed there, the distinction between [s] and [z] is fully categorical even when the same word alternates between [s]- and [z]-realizations. The distinction between the two phones is as clear as that between two distinct Italian phonemes -- for example, /f/ and /v/.

As long as the speakers fully produced the target word, their response was considered valid, even if they did not finish reading the carrier sentence. If the subjects corrected themselves while reading a sentence, the second reading was considered the valid one, since the interest here is in the performance of the subjects when they had understood the meaning of the target words, and not in mere “letter to sound” translation.

Three words beginning with the sequence bis were eliminated from the list, since they have a natural interpretation in which they are not analyzed as the prefix bi- followed by a stem beginning with /S/, but rather as the prefix bis- followed by a stem beginning with a vowel.

The words risurrezione and risuscitare were eliminated from the list because in most cases the speakers produced the synonym resurrezione instead of risurrezione and the synonym resuscitare instead of risuscitare. Thus, there were not enough correct responses to analyze these forms.

In the original list, there were two instances of the word risentito, one with the meaning ‘negatively affected (by something)’, and the other with the meaning ‘offended (by something)’. Only the entry with the more common former meaning was kept, since the frequency dictionary used for the statistical analysis (De Mauro, Mancini, Vedovelli & Voghera 1993) did not distinguish between the two meanings, and thus all the frequency-related factors had to be computed as if there was only one form. In any case, the two meanings are quite close, and the results for the two forms were similar.

After these trimmings, the list of valid forms was reduced to 102 target words plus 10 controls. The 102 target words with glosses are listed in the appendix.

3.2 On the reliability of the results

Two forms of control data indicate that the distribution of [s] and [z] emerging from the survey reflects the actual distribution of the two allophones in Northern Italian, and it is not the product of task-specific factors. First, all the controls with monomorphemic intervocalic /S/ were consistently produced with [z] by all speakers; all the controls with word-initial /S/ were constantly produced with [s] by all speakers (all the corresponding target words with V+SV displayed some variability between [z] and [s]). Thus, at least in the cases in which

the allophone choice is entirely phonologically driven, the speakers in the survey followed what we know to be the regular Northern Italian patterns.

Furthermore, we conducted a smaller scale survey in which 5 Northern Italian speakers were asked to judge the acceptability of [s] and [z] in the same words tested in the reading survey. While reading aloud and judging the acceptability of phonological strings are very different tasks, the results of the two surveys are remarkably similar.

The same 102 target word and 10 control word/phrase list used in the reading survey was part of the judgment survey corpus.⁶ Furthermore, the judgment survey corpus included 40 more control words: 20 with morpheme-internal intervocalic /S/ and 20 with word-initial /S/ before a vowel (and preceded in the carrier sentence by a word ending in vowel). Thus, there was a total of 102 target words with /S/ in a potential stem-initial context, and 50 controls: 25 control items with /S/ in morpheme-internal intervocalic context; 25 control items with /S/ in word-initial position.

The target and control items were presented on a computer screen. Each item was simultaneously presented in isolation and within a carrier sentence. For the 102 targets and 10 controls that also appeared in the reading survey, the carrier sentence was the same as in the reading survey. I decided to present the words in carrier sentences for two reasons: First, in some cases a carrier sentence is needed to disambiguate the intended meaning of a word. Second, I noticed in preliminary informal sessions with native Northern Italian speakers that presenting words in carrier sentences made the judgment task easier.

Five judges took part in the survey (none of them had been a subject in the reading survey). Each of them was asked to judge, for each word in the corpus, whether it was acceptable with both [s] and [z], only with [s] or only with [z]. Since the two allophones are spelled in the same way (<s>), [s] and [z] were defined as “the ‘s’ in ‘santo’ [= saint]” and “the ‘s’ in ‘casa’ [= house]”, respectively.

The 152 items were presented in a different random order to each judge, and the data collection was preceded by a short training session (4 items). The judges were asked

to press '3' on the computer keyboard for words only acceptable with [s]; '6' for words acceptable with both [s] and [z]; '9' for words only acceptable with [z]. Each item stayed on the screen until the judge made his/her choice by pressing a key, and the judges were given unlimited time to judge each form. After they pressed a key, the following item appeared on the screen after a 500 ms. interval. The judges were permitted to read the words and sentences aloud. After the data collection session, the judges were interviewed to make sure that their idiolect respected the Northern Italian distribution of [s] and [z].

The first step in the analysis of the data was to verify the reliability of the judges, by checking that they consistently judged morpheme-internal intervocalic /S/ in control forms to be acceptable only with [z], and word-initial /S/ to be acceptable only with [s]. Two judges mis-categorized a certain number of control forms, and consequently their data were discarded. The other three judges correctly categorized the /S/s of all the controls.

I coded the data of each of these three judges assigning the value 0 to words that were judged acceptable only with [z]; the value 1 to words that were judged acceptable with both allophones; and the value 2 to words that were judged acceptable only with [s]. Since the response patterns of the three judges were strongly correlated (the Spearman correlation coefficients computed pairwise for each pair of judges were higher than .7), their responses were combined into the compound variable JUDGES, ranging over 7 values, from 0 (only [z] acceptable for all judges), to 6 (only [s] acceptable for all judges). The JUDGES values of the target words are listed in the appendix.

JUDGES was then compared to N(s), which is the total number of [s] realizations per word produced by the 10 reading survey speakers whose response patterns are strongly correlated (see the discussion in section 4.3). The Pearson correlation coefficient between the two variables is higher than .9.

The high correlation between JUDGES and N(s) shows that the data on the distribution of IVB obtained by collecting acceptability judgments are similar to the data obtained through the reading survey. Since the judgment task is different from the reading

task in many respects (first and most importantly, judging and reading aloud are obviously different cognitive processes; furthermore, the judges, unlike the readers, did not have to perform their task under time pressure, and they were presented with each form only once), the fact that the results of the two surveys are highly correlated suggests that these results are not the product of task specific factors, but rather reliable evidence of the distribution of IVB in the speech of Northern Italians.

4. The representation of prefixed forms in the Italian lexicon: evidence from IVB

In this section, I discuss the evidence on the representation of potentially prefixed forms in the Italian lexicon that emerged from the IVB survey. In 4.1 and 4.2 the discussion is based on qualitative observations on subsets of data. In 4.3 the survey data are analyzed using various statistical techniques, in an attempt to determine the factors that induce the speakers to treat forms as morphologically complex.

4.1 Morphological constituency in lexicalized forms

First of all, it is interesting to observe that not only words with a completely transparent morphological structure (i.e. words that could be assembled using productive word-formation rules) are treated as morphologically complex by the speakers. The survey results show that some forms in which the prefix-stem structure is lexicalized are treated by speakers as morphologically complex.

For example, all the forms in (14) are lexicalized (in the sense that they could not be assembled by a productive word formation rule), and yet the surveyed speakers produced

the first two words in (14) with [s] 60/60 times, and the third word with [s] 59/60 times, which means that they treated their /S/s as stem-initial, and applied IVB:

- | | | |
|------|---------------|--------------------------|
| (14) | ri+[s]iede | ‘she/he resides’ |
| | ri+[s]olutivo | ‘resolutive’ |
| | ri+[s]aputo | ‘widely (or well) known’ |

In contemporary Italian, ri- ‘re-’ must be attached to verbs, and the resulting form is a verb with an iterative meaning:

- | | | |
|------|---------------------------------|--------------|
| (15) | ri + V -> [ri + V] _V | ‘to V again’ |
|------|---------------------------------|--------------|

In ri[s]iede, ri- is attached to a verb (meaning ‘to sit’), but the resulting form does not mean ‘to sit again’.⁷ Similarly, saputo is the past participle of the verb sapere ‘to know’, but ri[s]aputo does not mean ‘known again’. In ri[s]olutivo, -solutivo cannot be synchronically derived from any verbal paradigm, and the resulting word does not have an iterative meaning.

The lexicalized forms in (14) still preserve some of the semantic features that are usually associated with their stem, and apparently this is sufficient for the speakers to treat them as morphologically complex (notice that a form can be lexicalized, in the sense that it could not be assembled by a productive word formation rule, but still be, to some degree, semantically transparent, in the sense that its meaning is related to the meaning of one or both its subparts). The statistical analyses that I present in 4.3 below show that, indeed, semantic transparency is the most important factor inducing speakers to treat lexicalized forms as morphologically complex.

But why do speakers treat some lexicalized forms as morphologically complex in the first place? Notice first of all that Italian prefixes do not carry inflectional or syntactic information -- they do not mark gender, syntactic category, inflectional classes, etc. Thus,

storage of lexicalized prefixed forms in a morphologically complex format does not serve a syntactic or morphological purpose.

Furthermore, it is not likely that the morphologically complex representation of lexicalized forms serves some sort of “economy of storage” purpose. One could argue that completely transparent complex words with a compositional meaning do not need to be stored in the lexicon, since they can be reconstructed “on line” by putting the relevant morphemes together.⁸ However, words with a lexicalized meaning cannot be formed on line.

Even if risiede ‘(he/she) resides’ is represented as the combination of the iterative prefix re- and the stem -siede ‘(he/she) sits’, the speaker must also store some form of lexical unit marking the idiosyncratic properties of the complex form, since the complex word does not mean ‘to sit again’. For example, a lexicalized prefixed word could be represented as a single entry with a non-compositional semantic representation at the level of abstract lexical storage, but be connected with its component prefix and stem units at the level of phonological output representations. It is possible that this kind of representation would still be more economical than the representation in which the word is represented as an autonomous unit at all levels.

However, as we will see in the next section, speakers oscillate between [s] and [z] when producing some potentially complex words. This is probably due to the fact that speakers oscillate between a morphologically complex and a morphologically simple representation of these words. When the complex representation is selected, /S/ is stem-initial and IVB applies; when the simple representation is selected, /S/ is word-internal and thus it undergoes intervocalic voicing. If this interpretation of the cases with variable /S/ realization is correct, then we must accept that the speakers store double representations of some words. But, independently of the theory of morphological representations that one assumes, the decision to store double representations of some words cannot be motivated by economy of storage considerations. Thus, a theory explaining the complex

representation of lexicalized forms in terms of an economy of storage principle cannot account for the cases in which the same form is sometimes treated as simple and sometimes as complex.

A more plausible explanation of the fact that some lexicalized forms with /S/ in stem-initial position are treated as complex (at least for the sake of IVB) is based on the following observation: the phonological output representation of a form where IVB could in principle apply has a special status, since the form has to be produced in a different way depending on whether it is treated as complex or simple. Thus, when speakers encode the phonological output representation of a form with the relevant structure, they are exceptionally sensitive to any cue indicating that the form is prefixed.

For example, when the speakers learn the form risaputo ‘widely known’, they probably store the form as a single unit with a non-compositional semantic representation in the storage lexicon, since the word has an idiosyncratic meaning. Still, there is an obvious relation between the lexicalized meaning of risaputo ‘widely known’ and the meaning of the stem saputo ‘known’. The speakers are likely to be aware of this relation. Thus, they may choose to represent the word as ri+saputo in the phonological output lexicon, and consequently they apply IVB producing the /S/ in this form as [s].

If a prefixed word has a lexicalized meaning or structure, a morphologically complex representation of the word in the abstract lexical entry in which its meaning and grammatical properties are stored does not serve any function. Still, if the word is not completely opaque, speakers are likely to be aware of its (lexicalized) morphological structure and, if the word contains /S/ in stem-initial position, this may be sufficient for the word to be stored in a complex format in the phonological output lexicon.

In other words, according to the hypothesis presented here, the speakers decide how to represent words in the speech production lexicon independently of the way in which they represent the same words in the storage lexicon. In particular, speakers may be more inclined to represent a lexicalized word as complex in the phonological lexicon if the

word is subject to some morphophonological rule or constraint depending on its morphological status, such as IVB.

According to this hypothesis, then, only lexicalized forms in which IVB (or some other morphology-sensitive phonological process) applies may be treated as complex by the speakers, and only at the phonological output level. Further research should test the prediction that morphologically opaque forms are more likely to be represented as complex in the production lexicon if the application of some morphophonological rule/constraint depends on their morphological status.

Roelofs & Baayen (1996) present data on the representation of Dutch complex words in the production lexicon. On the basis of evidence from phonological and inflectional patterns sensitive to derivational morphology and results of experiments employing the implicit priming paradigm, they conclude that Dutch speakers treat some opaque forms as complex for the sake of speech production. Roelofs and Baayen interpret this result by claiming that “[i]n production, the memory organization and encoding algorithm aim at getting the forms right, and do not care about semantic transparency”. This view is similar to the one I am defending here.

4.2 “Ambiguous” morphology and variability in IVB

Since Northern Italian speakers are inclined to represent some lexicalized forms as morphologically complex, at least for the purpose of IVB, it becomes hard for them to decide whether a form is complex or not: they cannot simply check whether the form obeys the relevant productive word formation rule. A natural consequence of this situation is that speakers do not necessarily agree on the morphological analysis of the same words, and, more interestingly, even the same speaker can assign different morphological analyses to

the same word at different times. Thus, the distribution of IVB is subject to considerable variability within and between speakers. Consider for example the form in (16):

(16) [ko+'seno] / [ko'zeno] 'cosine'

Of the 12 speakers in the reading survey, 4 produced this word with [s] 5 times out of 5; 2 produced it with [z] 5 times out of 5; 2 produced [s] 4 times and [z] 1 time; 2 produced [s] 3 times and [z] 2 times; 2 produced [s] 1 time and [z] 4 times.

One could think that variability in a case like this is due to the fact that the application of IVB is optional. Under this view, /S/ in the form in (16) is always stem-initial but, since IVB is an optional process, sometimes it takes place and sometimes it does not.

However, if IVB were an optional process, then all the forms with the relevant prefix-stem structure should be produced sometimes with [s] and sometimes with [z]. This is not the case: certain forms (for example, asociale 'anti-social') are always produced with [s], whereas other forms (such as coseno) are sometimes produced with [s] and sometimes with [z].

Rather, the likely reason why forms like coseno are sometimes produced with [s] and sometimes with [z] is related to their morphologically "ambiguous" status. The prefix-stem structure of a word like coseno (or that of its English gloss 'cosine') is not as transparent as the one of a nonce formation such as risillabificare (or 're-syllabify'): one could not reconstruct the meaning of the complex form coseno from the meaning of its parts. Still, the structure of coseno is more transparent than the one of a word like presunto 'presumed' (which, indeed, is always produced with [z]). Hence, speakers oscillate between treating coseno as prefixed and monomorphemic. When they treat it as a prefixed form, /S/ is stem-initial and IVB must be respected. When they treat it as monomorphemic, /S/ is intervocalic within one morpheme, and intervocalic voicing is not blocked.

Interestingly, not all the forms with variable /S/ realization are like coseno, where [s] and [z] are equally likely to occur. Along with forms like coseno, where the number of [s] realizations is close to half the total number of tokens produced by all speakers in the survey (17.b), there are also forms such as the ones in (17.a), where [s] is far more frequent than [z], and forms like the ones in (17.c), where [z] is far more frequent than [s]:

(17)				<u>N of [s] realizations</u>
a.	resuscitare	‘to resurrect’		55/60
	risolvere	‘to solve’		53/60
	riservato	‘reserved’		52/60
b.	coseno	‘cosine’	36/60	
	residuo	‘(war) surplus’		31/60
	bisettrice	‘bisecting (line)’		28/60
c.	risalto	‘prominence’		16/60
	risarcire	‘to refund’		12/60
	desinenza	‘(morphological) ending’		2/60

The frequency of [s] realizations seems to depend on how “transparent” the morphological constituency of a word is. At one extreme, nonce formations must always be treated as morphologically complex, since speakers can only parse them if they are able to decompose them into morphemes. For example, if an Italian speaker hears the nonce formation risorpassare ‘to re-surpass’, he/she will have to access the prefix ri- and the verb sorpassare ‘to surpass’ in order to understand the new word. Thus, the speaker must be aware of the fact that risorpassare is complex. Indeed, all the native Northern Italian speakers I consulted have a very strong intuition that this form can only be produced with [s], i.e. treating /S/ as stem-initial and applying IVB.

At the other extreme, the word risotto, which is the name of a Northern Italian rice dish (etymologically, risotto is a diminutive form of riso ‘rice’), begins with the string ri-, but it is clearly not a prefixed form, given that its meaning is completely unrelated to the meaning of ri-, and given that -sotto is not a verb nor a deverbal noun, while ri- can only be attached to verbs or to names deriving from verbs. Indeed, all the native speakers that I

consulted have a very strong intuition that this form can only be produced with [z], i.e. treating /S/ as intervocalic within one morpheme.

Speakers can also be confident that very transparent lexicalized forms are prefixed and they do not analyze extremely opaque forms as synchronically prefixed. For example, risommergere ‘to resubmerge’ is not a nonce formation (risommergere is not uncommon, especially in phrases like risommergere uno scandalo ‘to hide a scandal’). Still, its meaning is perfectly predictable from the meaning of its parts and from the productive ri- prefixation rule. Thus, speakers do not have any problem in recognizing risommergere as a complex form, as shown by the fact that this form was always produced with [s] in the survey. The form risultato ‘result’, although diachronically prefixed, is completely opaque in synchrony, and speakers treat it as monomorphemic, as confirmed by the fact that it was always produced with [z] in the survey.

However, between such extreme examples, there are many cases in which it is not very clear whether a word is composed of a prefix and a stem or not. This fact is reflected in the behavior of speakers in terms of IVB: the harder it is to determine whether a form is prefixed or not, the more variation between [s] and [z] the form will involve.

Consider a case such as resuscitare ‘to resurrect’. This form cannot be constructed using productive morphology, since re- is not a productive prefix of Italian,⁹ and since the meaning of the form is lexicalized: suscitare means ‘to arouse’, but ‘to resurrect’ is not the same as ‘to arouse again’. Still, the semantic leap from ‘to arouse again’ to ‘to resurrect’ is very small, and thus speakers will tend to treat this form as prefixed. This is confirmed by the survey results: the speakers produced resuscitare with IVB 55 times out of 60.

Towards the other end of the continuum, we find forms like risalto ‘prominence’. The stem salto as an autonomous word means ‘jump’, a meaning which is not very close to the one of the complex form. Still, the composition of risalto is not totally opaque, in the sense that something prominent is something that jumps out from the background (cf. the English verb ‘to stand out’). Overall, speakers are likely to be more inclined to treat this

form as simplex, even if they may occasionally treat it as prefixed. This is confirmed by the survey data: the speakers produced the /S/ of risalto as [z] in the large majority of cases (44/60), but not always.

Finally, words like coseno fall in the middle of the continuum, and speakers have an equal chance of treating them as complex or simple. Again, this is reflected in the fact that similar words are produced with [s] or [z] with equal frequency.

To conclude, the variability of IVB application in lexicalized forms appears to be a consequence of the fact that the notion of morphological complexity relevant to IVB is a gradient one. This, in turn, is a natural consequence of the speakers' tendency to treat some lexicalized forms as morphologically complex: once the strict criterion that only forms which respect synchronic word-formation rules are represented as complex is abandoned, there is probably no clear-cut criterion that the speakers can use to distinguish between lexicalized complex and simple forms (see section 4.3 for a statistical analysis trying to determine which criteria affect the speakers' decision to treat some forms as complex).

In the previous section, we suggested that lexicalized forms in which IVB applies are represented as complex in the production lexicon, even if they are represented as single units in the storage lexicon. This approach can be easily extended to account for forms with variable /S/ realization: Speakers store two representations of such forms in the production lexicon, both linked to the same unit in the abstract storage lexicon. In one of the two output representations the form is stored in complex format, as a prefix + stem combination with stem-initial /S/. In the other representation the form is stored as a single unit, with monomorphemic intervocalic /S/. When the first representation is selected, the word is produced with [s]; when the second representation is selected, the word is produced with [z].

Still, we need to account for the fact that the likelihood of IVB application is not equal across forms with variable /S/ realization. Borrowing a common idea from the literature on lexical access (see Lively, Pisoni & Goldinger 1994 for a review), I

hypothesize that each of the two representations of a “morphologically ambiguous” word in the production lexicon is associated with a certain activation threshold. When the speaker wants to produce an ambiguous word, both its representations are activated, and the first one that reaches its activation threshold is the one chosen for production. The activation threshold of the two representations of a form is determined by the degree of confidence that the speaker has in the fact that the form is complex, plus some degree of (random?) transient fluctuation. For example, speakers are rather confident that resuscitare is a complex form, and thus they assign a much lower activation threshold to its complex representation than to the simple one. As a result, this word is usually produced with intervocalic [s]. However, there will also be rare cases in which, due to transient fluctuations, the activation threshold of the complex representation is unusually high and/or the activation threshold of the simple representation is unusually low. In such cases, the complex representation reaches its threshold first, and it is chosen for production.

In a case such as risalto, the speakers are not very confident that the form is complex, and thus assign a lower activation threshold to the non-prefixed representation of the word. As a result, this form is usually produced with [z].

In cases such as coseno, speakers are not more confident one way or the other, and thus they assign similar activation thresholds to the complex and simple representations. Consequently, the choice of the form to produce will depend entirely on the transient fluctuations in the threshold values. Thus, for truly ambiguous forms the [s] and [z] realizations are equally likely.

In order to make this model more explicit, we should develop a theory of what causes the fluctuations that produce temporary changes in the values of the activation thresholds. I will not deal with this issue here.

4.3 Factors determining morphological complexity

In the previous sections I showed that speakers treat some lexicalized forms as morphologically complex, and that (the relevant notion of) morphological complexity is a gradient one. At this point, it is natural to ask the following questions: Which factors induce speakers to treat some lexicalized forms as morphologically complex? Which factors make a morphologically ambiguous form more transparent or more opaque? In this section, I analyze the IVB survey data statistically, trying to assess the effect of several factors on the likelihood that a lexicalized form is treated as morphologically complex. Notice that all the words in the survey corpus are “lexicalized”, at least in the weak sense that they are not newly formed complex forms.

One may wonder why, instead of trying to assess the effect of a large number of factors by analyzing the survey results post hoc, I did not collect further data using factorial designs and testing the effect of a lower number of factors in more controlled settings. Unfortunately, this was not a viable strategy. As Laudanna & Burani (1995) and Smith (1988) have pointed out, it is difficult to construct factorial designs for the purpose of morphological processing experiments, since the potentially relevant factors are too many and they tend to be highly correlated with each other. In order to apply the IVB test, we are restricted to the few Italian forms that have the right morphophonological shape, and thus it is simply impossible to design balanced experiments.

The dependent variable used in the statistical analysis is the sum of [s] realizations per word produced by the 10 speakers whose response patterns were strongly correlated (the correlation coefficients computed pairwise for each pair of speakers were higher than .7). I refer to this variable as N(s). The N(s) values of the corpus words are listed in the appendix.

Given that speakers produce intervocalic [s] when they are treating a word as morphologically complex, the N(s) of a form is a measure of the probability that speakers

treat that form as morphologically complex. In the remainder of the section, I will present and discuss the factors that were computed for the purpose of the statistical analysis of the survey results:

Word Length (in Segments)

Word Length (in Syllables)

Stem Length (in Segments)

Stem Length (in Syllables)

If a word beginning with a prefix-like sequence is particularly long, there are more chances that the word is morphologically complex, since prefixed words are formed by a prefix plus another word (at least etymologically), and thus in general prefixed words will be longer than non-prefixed words. Also, it is possible that speakers are sensitive to the length of the string that follows the (potential) prefix: is the string long enough to be a plausible word/stem? This is what is measured by the Stem Length variables (here and after, the stem is the string formed by the root and all of the suffixes following it; in other words, the stem is the word minus the prefix).

Prefix Length (in Segments)

Prefix Length (in Syllables)

Root Length (in Segments)

Root Length (in Syllables)

Longer prefixes and roots are more “salient” (Laudanna & Burani 1995), in the sense that the probability of a long string occurring by chance is lower than the probability of a short string occurring by chance. Thus, speakers could be more confident in deciding that a word beginning with, say, para- is actually prefixed than in deciding that a word beginning with, say, a- is prefixed. Here and after, the root is the part of the word which remains when all prefixes and suffixes are stripped off.

Word Frequency

If a prefixed word is frequently used, it is more likely to acquire idiosyncratic properties, and hence to become morphologically opaque. On the other hand, new or recent morphological formations must be low frequency words.

Notice that no word potentially triggering IVB is extremely frequent. The most frequent word in the corpus is risolvere 'to solve', which is ranked as the 507th most frequent Italian word in the Lessico di Frequenza dell'Italiano Parlato (LIP).¹⁰

Prefix Frequency (Type)

Prefix Frequency (Token)

If a prefix is very common (either in the sense that it occurs in a number of different words, or in the sense that it occurs in very frequent words), the speakers may be more likely to assume that any word beginning with the phonological string corresponding to the prefix is prefixed. To compute the type frequency of a prefix, the number of LIP entries beginning with the prefix was counted.

A word was considered as beginning with a prefix if it met the following criteria: the word had to begin with the phonological string corresponding to the prefix; the main stress had to fall on the stem (in productively formed prefixed words, the prefix never carries main stress); the stem had to begin with a legitimate word-initial phonological string (this is another general feature of Italian prefixation); one of the meanings attributed to the prefix by Dardano (1978) and/or Zingarelli (1983) had to be somehow reflected in the meaning of the complex word. Token frequency was computed by adding the frequencies of all the LIP entries beginning with the relevant prefix.

Prefix Frequency / Pseudo-Prefix Frequency (Type)

Prefix Frequency / Pseudo-Prefix Frequency (Token)

Speakers may be more inclined to treat a word as prefixed if the word begins with a prefix-like string which mostly occurs in actually prefixed forms, independently of the absolute type frequency of the relevant prefix. Contrariwise, if an initial string corresponding to a prefix occurs in a large number of words that are not prefixed, speakers may be less inclined to treat words beginning with that string as prefixed.

The Prefix Frequency / Pseudo-Prefix Frequency value of a prefix is obtained by dividing the prefix frequency of the prefix by the frequency of pseudo-prefixed words beginning with a phonological string corresponding to the prefix. Pseudo-prefixed forms are forms that begin with a phonological string identical to a prefix, but are not actually prefixed: for example, industrial is a pseudo-prefixed form with respect to the English prefix in-. To be counted as pseudo-prefixed, a word has to meet the following criteria: it must begin with the phonological string corresponding to the prefix; the main stress must fall on the “pseudo-stem”; the pseudo-stem must begin with a possible Italian word-onset sequence; the word meaning is not related in any way to one of the meanings associated with the prefix.

Root Frequency (Productively Related Forms Only) (Type)

Root Frequency (Productively Related Forms Only) (Token)

Root Frequency (Type)

Root Frequency (Token)

As with prefixes, it could be the case that speakers are more confident about treating a potential root as an actual root if the root is very common, in terms of type or token frequency.

For each word in the corpus a list of the words containing the same root was prepared using Gabrielli (1996), an Italian dictionary on CD-ROM containing 300,000

forms. Words with the same prefix of the corpus word were not counted (while the existence of the word permit may lead the speaker to postulate a root -mit in submit, there is no reason why the word submitted should count as evidence leading the speaker to treat submit as sub + mit). Words not familiar to the author or words that could be considered morphologically complex nonce formations (such as to re-stain or to re-die) were also excluded. For each paradigm of words sharing the same root, the total token frequency was calculated using the LIP.

It is not always easy to decide whether two words share the same root or not. First of all, we have to deal with a phonological issue: should only words that share an identical phonetic string be counted as forms with a common root, or can also words containing different allomorphs of the same morpheme be counted as sharing the same root? I adopted a conservative phonological criterion: only words sharing segmentally identical roots were treated as related. Differences in stress placement were tolerated, since stress shift is a common consequence of many productive inflectional and derivational suffixation processes of Italian. Obviously, differences in the voicing value of stem-initial /S/ were also ignored.

The reason for adopting a conservative phonological criterion is simply that it is very easy to check whether two words share segmentally identical roots. On the other hand, if we accept that different phonological strings can be instances of the same root, it becomes extremely hard to determine which words share the same roots on the basis of straightforward operational criteria. For example, do permit and permission share the same root? What about semantics and sign?

Second, we have to decide whether it is sufficient for two words to share a phonological string and the -- often vague -- meaning associated with that string to conclude that those two words share the same root, or whether the two words should also be part of some systematic morphological paradigm, such that, given one word, we can derive the other. For example, the words persist and insist share the phonological string -

sist, and the two words appear to be semantically related. However, we could never construct insist on the basis of persist (or vice versa) using the rules of synchronic English morphology. On the other hand, if an English speaker knows the word uncool, she/he can understand/construct the word ultra-cool on the basis of the generalization that, if un-X means ‘not X’, then ultra-X means ‘very X’. On one hand, it is not completely unreasonable to assume that insist and persist share the same root; on the other hand, it is clear that the relationship between uncool and ultra-cool is stricter and more transparent. Thus, I computed two pairs of Root Frequency factors: Root Frequency (Productively Related Forms Only) (Type) and Root Frequency (Productively Related Forms Only) (Token) are computed counting only forms that are related by systematic and productive patterns of Italian morphology, whereas Root Frequency (Type) and Root Frequency (Token) also count words in which the root can be identified on phonological and semantic grounds, even if their relationship with the target word is not productive.

Root Frequency (Productively Related Forms Only) / Pseudo-Root Frequency (Type)

Root Frequency (Productively Related Forms Only) / Pseudo-Root Frequency (Token)

Root Frequency / Pseudo-Root Frequency (Type)

Root Frequency / Pseudo-Root Frequency (Token)

As with prefixes, chances are that the speakers will be more confident that a certain word is complex if its potential root is a phonological string that mainly occurs in forms in which it functions as a real root. A word containing a phonological string identical to a root contains a pseudo-root if the string is either word-initial, or if it occurs after one or more prefixes or pseudo-prefixes and the meaning of the word does not reflect one of the possible meanings associated with the root (e.g., personal contains the pseudo-root son- (cf. resonate)). Pseudo-roots were searched on the Gabrielli (1996) CD-ROM (token frequency values were calculated using the LIP).

Prefix Transparency

Stem Transparency

Semantic transparency is likely to play a very important role in the decision to treat a form as morphologically complex. If the meaning of a potentially complex form contains some subpart identifiable as the meaning of the prefix, and some subpart identifiable as the meaning of the potential stem, then the form is likely to be prefixed.

Notice that here I use the term “semantic transparency” to refer to something different from “semantic compositionality”. A form is semantically compositional if its meaning is entirely predictable from the meaning of the parts and from the rules that assemble those parts. A form is semantically transparent if it is possible to establish some relationship between its meaning and the meaning of its subparts. Semantic compositionality entails semantic transparency, but not vice versa.

Semantic transparency is often a matter of vague sound-to-meaning correspondences. For example, the Italian speakers that I consulted share the intuition that the stem of the word antesignano ‘forerunner’ is not entirely semantically opaque, even if the root sign- ([ˈsiɲ]) only appears in other lexicalized forms with a rather different meaning, such as significato ‘meaning’. The reason why sign- is not considered entirely opaque is probably that it is phonologically similar to the etymologically related form segn- ([ˈseɲ]), the root of segno ‘sign’. In a way, a forerunner is a beforehand “sign” of the future birth of a certain movement. Semantic transparency, unlike compositionality, is a gradient property: antesignano is not as transparent as, say, risillabificare ‘resyllabify’ but more transparent than, say, deserto ‘desert’.

It is hard, if not impossible, to compute semantic transparency values mechanically. Thus, I asked the same speakers that took part in the reading survey to rate the degree of semantic transparency of each word of the corpus on a four point scale.¹¹ Since it is not necessarily the case that prefixes and stems are equally transparent in semantic terms, I asked the speakers to rate the degree of transparency of prefixes and stems separately. The

speakers were given an unlimited amount of time to fill out the semantic transparency questionnaire.

The instructions provided an example of a word in which both the prefix and the stem are completely transparent (the word rileggere 'to re-read'), and an example of a word in which both the prefix and the stem are totally opaque (the word ricevere 'to receive') Subjects were asked to rate the forms on two four point scales, treating rileggere as an example of a word which should be assigned a 4-4 rating, and ricevere as an example of a word which should be assigned a 1-1 rating.

The ratings of the 8 speakers that returned the form were compared in a correlation matrix. The prefix transparency ratings of two speakers were poorly correlated with those of the others (the correlation coefficient in pairwise comparisons of the scores of each of these two speakers and some of the others was lower than .5). The prefix transparency ratings of these speakers were not considered. The stem transparency ratings of one of these speakers were poorly correlated with those of the others (correlation coefficient < .5 in some pairwise comparisons). The stem transparency ratings of this speaker were not considered. Consequently, Prefix Transparency ratings are averaged across 6 speakers; Stem Transparency ratings are averaged across 7 speakers.

I decided to collect semantic transparency ratings from the same speakers who participated in the survey since lexico-semantic ratings are likely to depend in part on speaker-specific variables, such as knowledge of Latin or Ancient Greek, or familiarity with specialized lexica. Thus, semantic transparency values computed on the basis of judgments from the same speakers are probably more adequate as predictors of the speakers' behavior. At the same time, the fact that the semantic transparency ratings are averaged across speakers guarantees that speaker-specific biases are minimized.

Stem AutonomyAutonomous Stem Frequency (Token)Autonomous Stem Frequency / Word Frequency (Token)Word-Initial RootWord-Initial Root Frequency (Type)Word-Initial Root Frequency (Token)

Kenstowicz (1995) has proposed that IVB is a paradigmatic effect. According to Kenstowicz's theory, stem-initial /S/s in prefixed words are produced as [s] by analogy with the corresponding non-prefixed words, where /S/ is word initial, and consequently it is produced as [s]. For example, the word a[s]lociale 'anti-social' is produced with [s] by analogy with its base, the word sociale 'social', in which /S/ is word initial, and thus realized as [s].

While the major strength of Kenstowicz's approach is that it naturally derives the asymmetry between the stem-initial position and the morphological contexts in which IVB does not occur, his theory is also making a strong prediction about the distribution of IVB among words with stem-initial /S/. According to the paradigmatic analogy theory, only prefixed words with a free stem, i.e. a stem which also occurs as an independent word, will be subject to IVB. If the stem of a prefixed form does not occur as an independent word with initial [s], then there is no base form triggering an analogic effect, and intervocalic voicing should not be blocked. Thus, Kenstowicz can elegantly explain the fact that a form such as resistente 'resistant' is always produced with [z]. The stem -sistente is not an independently existing word, and hence the /S/ of resistente is free from any analogic pressure, and, being in intervocalic position, it is realized as [z].

The data from our survey falsify Kenstowicz's prediction: there are several words whose stem is not an independent word, but which are nevertheless (often or always) subject to IVB.¹² For example:

(18)		<u>N of [s] realizations</u>	
	antesignano	'forerunner'	47/60
	asintoto	'asymptote'	59/60
	polisemo	'polysemous'	54/60
	polisindeto	'polysindeton'	60/60
	presidio	'garrison'	54/60
			*signano
			*sintoto
			*semo
			*sindeto
			*sidio

Thus, the distribution of IVB cannot be entirely explained in terms of paradigmatic analogy. However, it is still possible that paradigmatic analogy is one of the factors making IVB more or less likely. To test this possibility, we computed the Stem Autonomy variables and the other variables in this set.

Stem Autonomy is a binary variable. If the stem of a form exists as an independent word, the form is assigned a Stem Autonomy value of 1. Otherwise, the form is assigned a value of 0.

The Autonomous Stem Frequency variable was computed to test whether analogic effects depend on the frequency of the stem as an autonomous word. For example, if the stem occurs as an autonomous word very frequently, the analogic pressure may be stronger.

The Autonomous Stem Frequency / Word Frequency factor could be relevant if what matters is the ratio of the frequency of the stem as an autonomous word over the frequency of the complex word itself. In principle, the analogic pressure could be stronger in the cases in which the stem as an autonomous word is more frequent than the complex word

Even if the stem as such is not an independent word, it is possible that, as long as its root occurs word-initially, analogic effects favoring [s] are active. This justifies the binary factor Word-Initial Root and the two frequency-based factors Word-Initial Root Frequency (Type) and Word-Initial Root Frequency (Token).

Pre-Stress /S/

Post-Stress /S/

The rationale for the binary Pre-Stress /S/ and Post-Stress /S/ variables is that /S/s that are adjacent to the stressed vowel could in principle behave differently from other /S/s due to their special prosodic position. For example, a pre-stress /S/, being the onset of the stressed syllable, could be longer, which would make an [s]-realization more likely. The Pre-Stress /S/ factor sorts out words in which /S/ is adjacent to the stressed vowel, occurring to its left; Post-Stress /S/ sorts out words in which /S/ is adjacent to the stressed vowel, occurring to its right.

Notice that the last two sets of variables differ from the previous ones, in that they measure phonological rather than morphological effects (surface analogy to other forms in the same paradigm and stress location).

4.3.1 Statistical analysis

It was impossible to assess the effect of each independent variable on N(s) in a single multiple regression, because there are too many independent variables, and many of these are strongly correlated with each other. In order to reduce the number of independent variables to enter in the multiple regression, I adopted the following strategy.

I first selected the set of independent variables that were significantly correlated with N(s) in both a Pearson and a Spearman correlation analysis with $\alpha = .05$.¹³ For each cluster of highly correlated variables in this set, I selected the variable with the highest Pearson correlation coefficient with N(s). This left me with a subset of 10 independent variables.

Since the collinearity indices for this subset were very low, I proceeded with a stepwise multiple regression including the 10 independent variables and their interactions,

with N(s) as the dependent variable. The only factors entered in the stepwise procedure were Stem Transparency ($p < .00$) and Prefix Length (in Syllables) ($p < .03$). The other factors and the interactions had no significant effect on N(s). The r^2 value of the regression model is .416.

4.3.2 Discussion

Semantic transparency (and the transparency of the stem in particular)¹⁴ is the major factor affecting the speakers' choice to treat a form as morphologically complex. Prefix Length (in Syllables) is also a significant predictor of N(s).

The importance of semantic transparency is not surprising: if a word can be parsed into a prefix-like string and a stem-like string, the fact that aspects of the meaning of the potential prefix and/or stem are reflected in the whole word's meaning is likely to make the speakers more confident in treating the word as morphologically complex. As noted above (see discussion in section 4.3), semantic transparency refers here to something different from semantic compositionality. Semantic transparency is often a matter of vague sound-to-meaning correspondences, and it is a gradient property. Given that speakers rely on semantic transparency as the main heuristic criterion to determine the morphological status of words for the purpose of IVB, it naturally follows that the relevant notion of morphological complexity is a gradient one.

Prefix Length (in Syllables) is the other factor with a significant effect on N(s). The probability that a four or five segment string occurs by chance is much smaller than the probability that a two or three segment string occurs by chance. Thus, speakers will be more inclined to guess that a lexicalized word is prefixed if the word starts with a string corresponding to a long prefix. A striking example of the effect of Prefix Length (in Syllables) is the one of the word parasanghe [para+sange] 'parasangs (ancient Persian

measurement unit)'. This form is entirely opaque from a semantic point of view: indeed, parasanghe is not even a prefixed word from an etymological point of view. Nevertheless, since para- is a long prefix, the N(s) value of this form is relatively high (37/50).

To conclude the discussion of the statistical results, I must remark that the factors entered in the statistical analysis could only account for about 42% of the variation in N(s). Thus, while some of the variables computed have a significant effect on the distribution of IVB, they are far from entirely accounting for its distribution.

5. Conclusion

Typically, morphophonological studies focus on how morphological structures affect phonological processes. In this paper, I took the opposite approach, studying the phonological distribution determined by a morphophonological process as a source of evidence on the morphological representations posited by speakers in the phonological output lexicon. As long as a phonological pattern sensitive to morphological constituency is productive and systematic, as is the case for Northern Italian IVB, it is possible to study its distribution to gather evidence on the non-trivial issue of which forms are represented by the speakers of a certain language as morphologically complex (at least for the sake of morphophonological processes).

The evidence from IVB presented in this paper shows that Italian speakers represent some lexicalized forms (forms with idiosyncratic semantic and/or morphosyntactic properties) as morphologically complex, at least for the sake of this morphophonological process. I proposed to explain this fact in the following terms: When speakers encode the phonological output representation of a form in which IVB could in principle apply, they pay special attention to any cue possibly indicating that the form is prefixed, since the form has to be produced in different ways depending on its morphological status. If the

morphological constituency of a form is not completely opaque, speakers may be aware of it, and they may decide to store a complex representation of the form in the phonological output lexicon, even if the form, because of its idiosyncratic meaning, is represented as a single unit in the abstract lexical entry in which semantic and grammatical features are stored.

This hypothesis predicts that only lexicalized forms in which IVB (or some other morphophonological process) applies may be represented as complex by speakers, and only in the phonological output lexicon. Further research should test these predictions.

The data presented here also show that speakers tend to oscillate between complex and simple representations of some words with the relevant structure. This behavior is reflected in the fact that these words are sometimes produced with [s] and sometimes with [z]. The distribution of [s] and [z] realizations is not uniform across the words which display variability: words whose morphological structure is more “transparent” are produced with [s] (i.e. treated as morphologically complex) more often than words whose morphological structure is more “opaque”. Thus, it appears that the relevant notion of morphological complexity is a gradient one.

It is likely that this is a consequence of the tendency to store some lexicalized forms as morphologically complex: while it is possible to sort out forms respecting synchronic word-formation rules in a categorical, binary way, there is probably no clear-cut criterion determining which lexicalized forms should be represented as complex and which forms should be stored as single units.

I proposed that each word with variable /S/ realizations has two representations in the production lexicon, both linked to the same abstract unit. In one of the two representations, the word is stored as a prefix + stem combination with stem-initial /S/ (which is realized as [s]); in the other, the word is represented as a single unit with monomorphemic intervocalic /S/ (which is realized as [z]). I also presented an account of the unequal distribution of [s] and [z] realizations in words with variable IVB application,

based on the idea that the two output representations of a morphologically ambiguous word are associated with activation thresholds reflecting the speakers' degree of confidence in the fact that the word is complex.

A statistical analysis of the data showed that, not surprisingly, the major factor affecting the speakers' decision to treat a lexicalized form as morphologically complex is the semantic transparency of the form. However, a purely form-based criterion also affects the speakers' behavior: if a prefix is phonologically long, there is a higher chance that words beginning with a string identical to the prefix will be treated as prefixed.

*I wish to thank Donca Steriade, Ed Stabler, Sun-Ah Jun, Jody Kreiman, Sumiko Takayanagi, Janet Pierrehumbert, Motoko Ueyama, Adam Albright, the members of the UCLA phonetics laboratory, Pier Marco Bertinetto, Michele Loporcaro, Harald Baayen, the reviewers for the Yearbook of Morphology and, especially, Bruce Hayes and Carson Schütze for help and advice.

¹With the term Northern Italian, I refer to the variety of Standard Italian spoken by cultivated Northern Italian speakers.

²The data presented in this paper, unless otherwise specified, are drawn from the survey discussed in section 3, or based on unanimous judgments by the author and at least five other speakers of Northern Italian.

³Loporcaro (1996) formulates Intervocalic /S/ Voicing as a morpheme boundary sensitive rule. Nespor & Vogel (1986), Vogel (1994) and Peperkamp (1995) claim that IVB is a prosodic phenomenon, depending on the fact that intervocalic /S/ voicing only applies within prosodic words: according to these theories, a prosodic word boundary separates /S/ from the vowel preceding it in forms such as asociale. Still, even for these authors a and sociale are parsed into different prosodic constituents because they belong to different morphological constituents (a is a prefix and sociale is a stem). In the word a[z]ino 'donkey', a is not a prefix, and consequently it is parsed in the same prosodic word as [z]ino, and intervocalic voicing applies. Thus, although in this approach intervocalic voicing is a prosodic rule, the fact that words such as the ones in (2) are subject to IVB still depends on the fact that they are morphologically complex (since they are morphologically complex, their subparts are parsed into prosodic constituents whose boundaries block intervocalic voicing). Kenstowicz (1995) claims that IVB is a paradigm-driven analogic effect: asociale is produced with [s] by analogy with the paradigmatically related word sociale, where /S/ is word-initial, and hence realized as [s]. Still, the reason why asociale is subject to analogic pressure from sociale is because asociale is a complex word morphologically related to sociale. Thus, even within the "analogic" approach IVB ultimately depends on morphological constituency.

⁴Here and below, when discussing potentially prefixed forms, I often refer to non-prefixed words as monomorphemic. Notice, however, that usually these non-prefixed words are not truly monomorphemic, since they bear, at least, an inflectional suffix (and, occasionally, one or more derivational suffixes). For example, when I claim that speakers treat presunto as a monomorphemic word, I mean that they do not treat it as a prefixed form -- but still it is likely that they are aware of the fact that the final -o is the masculine singular suffix. Thus, monomorphemic is used here as a synonymous of non-prefixed.

⁵The judgment survey described in section 3.2 is not based on a speech production task. Nevertheless, judges were giving metalinguistic judgments on the way in which they would produce words with the relevant structure. Given also the high correlation between the reading and judgment survey results, it is reasonable to assume that the judges based their responses on their intuitions about phonological output representations.

⁶The target items that were trimmed during the preliminary analysis of the reading survey results (see the discussion in section 3.1) were not included in the judgment survey corpus.

⁷The meaning 'to sit again' is possible if the word is interpreted as a nonce formation constructed by attaching ri- to siede. However, this reading is ruled out by the carrier sentence, in which only the

lexicalized reading 'to reside' is acceptable. The carrier sentence is: Gianni Bianchi risiede in via Roma 7 'Gianni Bianchi resides at number 7, Roma street'.

⁸Notice, however, that some experimental evidence suggests that speakers store even regularly inflected words in the lexicon (see for example Baayen, Dijkstra & Schreuder 1997 on regular Dutch plurals and Sereno & Jongman 1997 on regular English plurals).

⁹The prefix re- can only be used productively as the allomorph of ri- before stem-initial /i/.

¹⁰The LIP (De Mauro, Mancini, Vedovelli & Voghera 1993), a spoken Italian frequency dictionary, was used for the computation of all the frequency variables. Token frequency values are based on the actual number of occurrences of words (or prefixes/roots) in the LIP corpus, and not on frequency ranks. Word Frequency values represent the frequency of lexemes (= sets of inflectionally related forms constituting single LIP entries), and not the frequency of the specific inflected forms presented in the survey.

¹¹Because of the length and difficulty of this task, I did not add fillers.

¹²There are also prefixed forms which are not subject to IVB even if their stem is an independent word. For example, presentimento 'premonition' is more often produced with [z] than [s], even if the stem sentimento 'feeling' is an independent word. However, this and similar cases are not too problematic for Kenstowicz's theory, since one could argue that synchronically presentimento and sentimento are no longer paradigmatically related, and hence the former is not subject to analogic pressure from the latter. Of course, without an explicit theory of paradigmatic relatedness, an explanation along these lines is ad hoc.

¹³Logarithmic transformations of the token frequency variables were used in the Pearson correlation analyses.

¹⁴Prefix Transparency is also highly correlated with N(s) (the correlation coefficient between Prefix Transparency and N(s) is .44 in the Pearson analysis and .47 in the Spearman analysis). However, Prefix Transparency is highly correlated with Stem Transparency (the Pearson coefficient between the Transparency variables is .82), and the latter has a higher correlation with N(s) (.62 in the Pearson analysis, .63 in the Spearman analysis). For this reason, Prefix Transparency was not among the factors used in the regression analysis. The fact that the transparency of stems has a stronger effect on the likelihood of IVB than the transparency of prefixes is probably due to the fact that stems tend to have more specific meanings than prefixes and, thus, they are easier to identify on pure semantic grounds.

REFERENCES

- Baayen, Harald; Dijkstra, Ton; and Schreuder, Robert (1997). Singulars and plurals in Dutch: evidence for a parallel dual-route model. Journal of Memory and Language 37, 94-117.
- Baroni, Marco (1996). The allophonic distinction between [s] and [z] in Northern Italian, and what speakers know about it. Manuscript, UCLA.
- Baroni, Marco (1998). The phonetic nature of the Northern Italian allophones [s] and [z] in words with variable realization: electroglottographic and acoustic evidence. UCLA Working Papers in Phonetics 96, 166-174.
- Burani, Cristina (1992). Patterns of inflectional errors with reference to the Italian adjectival system. Rivista di Linguistica 4, 255-272.
- Canepari, Luciano (1992). Manuale di pronuncia italiana. Bologna: Zanichelli.
- Caramazza, Alfonso; Laudanna, Alessandro; and Romani, Cristina (1988). Lexical access and inflectional morphology. Cognition 28, 297-332.
- Dardano, Maurizio (1978). La formazione delle parole nell'italiano d'oggi. Roma: Bulzoni.
- De Mauro, Tullio; Mancini, Federico; Vedovelli, Massimo; and Voghera, Miriam (1993). Lessico di frequenza dell'italiano parlato. Milano: Fondazione IBM Italia.

Feldman, Laurie (ed.) (1995). Morphological aspects of language processing. Hillsdale: LEA.

Gabrielli, Aldo (1996). Il Gabrielli su CD-ROM (CD-ROM). Milano: Elemond.

Henderson, Leslie (1985). Towards a psychology of morphemes. In Progress in the Psychology of Language, vol. 1, Andrew Ellis (ed.), 15-72. London: LEA.

Kenstowicz, Michael (1995). Base-identity and uniform exponence: alternatives to cyclicity. Manuscript, MIT.

Laudanna, Alessandro; and Burani, Cristina (1995). Distributional properties of derivational affixes: implications for processing. In Feldman (1995), 345-364. Hillsdale: LEA.

Lively, Scott; Pisoni, David; and Goldinger, Stephen (1994). In Handbook of Psycholinguistics, Morton Ann Gernsbacher (ed.), 252-301. San Diego: Academic Press.

Loporcaro, Michele (1996). Prosodic domains in Romance phonology. Manuscript, University of Zürich.

Marslen-Wilson, William; Tyler, Komisarjesky Lorraine; Waksler, Rachelle; and Older, Lianne (1994). Morphology and meaning in the English mental lexicon. Psychological Review 101, 3-33.

Nespor, Marina; and Vogel, Irene (1986). Prosodic phonology. Dordrecht: Foris.

Peperkamp, Sharon (1995). Prosodic constraints in the derivational morphology of Italian.

Yearbook of Morphology 1994, 207-244.

Roelofs, Ardi; and Baayen, Harald (1996). Semantic transparency in producing polymorphemic words. Submitted.

Sandra, Dominiek; and Taft, Marcus (eds.) (1994). Morphological structure, lexical representation and lexical access (special issue of Language and Cognitive Processes). Hove: LEA.

Sereno, Joan; and Jongman, Allard (1997). Processing of English inflectional morphology. Memory & Cognition 25, 425-437.

Smith, Philip (1988). How to conduct experiments with morphologically complex words. Linguistics 26, 699-714.

Vogel, Irene (1994). Phonological interfaces in Italian, In LSRL 23, 109-126.

Zingarelli, Nicola (1983). Il nuovo Zingarelli. Bologna: Zanichelli.

Zingarelli, Nicola (1987). Il nuovo Zingarelli minore. Bologna: Zanichelli.

APPENDIX

The following table presents the 102 target words of the survey corpus with English glosses, the N(s) value of each word (see section 4.3), the Total N(s) value of each word (computed by summing the number of [s] responses of all the 12 speakers) and the JUDGES value of each word (see section 3.2).

Word	Gloss	N(s)	Total N(s)	JUDGES
antesignano	forerunner	41	47	6
antisettico	antiseptic	50	60	6
asepsi	asepsis	47	57	6
asessuato	asexual	50	60	6
asettico	aseptic	50	60	6
asimmetrico	asymmetrical	49	59	6
asincrono	asynchronous	50	60	6
asindeto	asyndeton	43	53	5
asintoto	asymptote	50	59	6
asintotica	asymptotic	49	59	6
asociale	antisocial	50	60	6
bisecante	bisecting	48	58	6
bisecolare	bisecular	50	60	6
bisessuale	bisexual	50	60	6
bisessuati	hermaphroditic	50	60	6
bisestile	referring to leap-year	1	1	0
bisesto	referring to leap-year	21	31	2
bisettimanale	bi-weekly	50	60	6
bisettrice	bisecting (line)	26	28	1
bisezione	bisection	48	58	5
bisillabo	disyllabic	50	60	6
bisolfito	bisulphite	49	59	6
coseno	cosine	31	36	4
desacralizzare	desacralize	50	60	6
deserto	desert	0	0	0
desiderio	desire	0	0	0
designato	designated	1	1	0
desinare	to dine	0	0	0
desinenza	grammatical ending	2	2	0
desistere	to desist	0	0	0
desolato	desolate	1	1	0
desueto	obsolete	14	14	0
desumere	to deduce	2	2	0

Word	Gloss	N(s)	Total N(s)	JUDGES
iposolfito	hyposulphite	50	60	6
metasemia	metasemy	50	56	6
parasanghe	parasangs	37	42	5
paraselene	paraselene	50	60	6
parasimpatico	parasympathetic	47	57	6
polisemo	polysemous	49	54	6
polisillabo	polysyllabic	50	60	6
polisindeto	polysyndeton	50	60	6
presagio	presage	0	0	0
presalario	stipend	50	60	6
preselezione	preliminary selection	49	59	6
presentimento	premonition	10	10	0
preservare	to preserve	44	49	3
preservativo	prophylactic	49	53	3
preside	(school) principal	49	58	2
presidente	president	38	43	1
presidiato	garrisoned	43	48	3
presidio	garrison	46	54	5
presiede	(he/she) presides	49	59	5
presuntivo	estimate	0	0	0
presunto	presumed	0	0	0
presuntuoso	presumptuous	0	0	0
presunzione	presumption	1	1	0
presupposizione	presupposition	1	1	2
prosecuzione	prosecution	48	58	4
proseguì	go on (imp)	50	60	5
proseguimento	continuation	48	57	5
residenza	residency	11	13	0
residenziale	residential	27	32	0
residuato	(war) surplus	26	31	1
resiliente	resilient	16	21	1
resipiscenza	resipiscence	23	32	2
resistente	resistant	0	0	0
Resistenza	Resistance	0	0	0
resisti	resist (imp)	0	0	0
resurrezione	resurrection	49	59	4
resuscitare	to resurrect	46	55	5
risacca	undertow	31	41	3
risalita	re-ascension	49	59	5
risalta	(it) stands out	15	15	0
risalto	prominence	16	16	0
risanabile	curable	41	51	5
risanamento	curing	41	51	5
risaputo	widely known	49	59	6
risarcibile	refundable	13	13	0
risarcire	to refund	12	12	0
risentimento	resentment	48	58	4

Word	Gloss	N(s)	Total N(s)	JUDGES
risentito	been negatively affected	45	53	5
riserva	reserve	50	55	3
riservatezza	circumspection	44	49	4
riservato	reserved	47	52	3
risiede	(he/she) resides	50	60	5
risollevalo	raised again	50	60	6
risolubile	resolvable	45	55	4
risolutivo	resolutive	50	60	4
risoluto	resolute	47	56	2
risoluzione	resolution	46	56	3
risolvere	to resolve	48	53	4
risommergere	to resubmerge	50	60	6
risonanza	resonance	46	56	5
Risorgimento	“Risorgimento”	47	57	5
risorgive	underground river springs	43	53	6
risorse	resources	50	56	5
risorto	resurrected	48	53	5
risucchio	eddy	47	57	6
risultato	result	0	0	0
risuona	(it) resonates	48	58	6
trasognata	dreamy	48	54	5
trasuda	(it) transudes	50	60	6