VOWEL INSERTION IN THE ITALIAN ADAPTATION OF ENGLISH LOANWORDS

Linguistic constraints and the role of L2 input

Research Master's Thesis in Linguistics

Amsterdam, June 2018

Veronica Miatto (11106263)

Supervisor:

Mw. Dr. Silke Hamann

Second reader:

Dhr. Prof. Dr. Paul Boersma

Abstract

Stochastic theoretical frameworks, such as stochastic OT (Boersma 1997) or Neural Networks (Boersma et al. 2013), argue for the establishment of speech categories from the statistics of the speech signal. The fundamental point of these models is that learning is achieved through the recognition of patterns due to repeated input. This means that categories or constraints can be modified or formed if enough input is received. The main aim of this research is to establish whether the production of vowel insertion after consonant-final English loanwords in Italian is influenced by the amount of native English input received through both passive and active interaction. Previous studies that have investigated the influence of the L2 in the adaptation of loanwords have measured knowledge in the L2 through L2 proficiency, age of arrival, and bilingualism (Flege et al. 2003: Kang & Schertz 2017; Kwon 2017; Nomura & Ishikawa 2018). A secondary aim is to confirm the effect of linguistic constraints, such as voicing and type of consonant, on the production of vowel insertion in the Italian adaptation of English loanwords (following Grice et al. 2015).

Keywords

L2 proficiency, epenthetic vowels, paragogic vowels

Acknowledgments

This thesis would not have been possible without the contribution of many people. First and foremost, I want to thank Silke Hamann, not only for the endless patience and guidance she offered me throughout the year, but also for supporting all my ideas, which has allowed me to grow and learn on many levels, and for all the advice and chats on academic life. I also want to thank Paul Boersma, for his ever-useful advice and all the challenging remarks from the first stages of the thesis. Having you both as supervisors during these two years has been invaluable and really shaped me as a (future) researcher.

I am grateful to the ISTC (Istituto di Scienze e Tecnologie della Cognizione) in Padua, Italy, for letting me use their phonetic laboratory for data collection, and for giving me a place to work during my stay. Pastries are really not enough as a payment. I am particularly grateful to Alberto Benin, for giving me all the equipment and the assistance that I needed in setting up the experiment, and to Cinzia Avesani, for helping me once again in my research, and taking the time to discuss the experiment and life as a researcher in general. Thanks also to all the participants of the experiments, and to everyone I tested the experiment with beforehand.

Without the support of my friends here, I would have succumbed to the combined stress of thesis and PhD applications. Thanks to Silvia, Catarina, Jorge and Felicia for the hospitality, food, acceptance, and for essentially being the reason why I am starting a PhD in the US. To all my classmates for being my companions in this hard journey and such a fun gang to hang out with. Thanks also to Francesca, Agnese and Lucia, who made my first year in Amsterdam truly special.

I had the opportunity to discuss the very first stages of this study with many people, who invariably gave me useful advice and papers to read. I am grateful to all of them, and in particular to Lori Repetti, Jennifer Nycz and Francisco Torreira.

Finally, this thesis truly would not exist without my family, who has always incessantly supported me on every possible level. You always encouraged me to develop my talents and pursue my dreams, and you gave me the means to do so. Thank you for your unwavering love, and for standing by me wherever my vocation is taking me.

Table of contents

1. INTRODUCTION

The present study tackles loanword adaptation and the influence that input from the source language can have in the production of said loans. Research in loanword adaptation has exploded in the last decade, as the mismatches between foreign phonological systems in perception and production can inform us on how the latter works in the human mind. Particular focus is here given to the insertion of vocalic elements in Italian after consonant-final loans, also called in the literature 'epenthetic vowels' (EP) or 'paragogic vowels'.

EPs are usually thought of as a repair strategy for a violated syllabic structure in the borrowing language, although there is some controversy as to what exactly motivates their insertion (Hall 2011). EPs in Italian have been investigated by several authors from different perspectives. For instance, Cardinaletti & Repetti (2007) investigated the phenomenon from a morphophonological point of view. They stated that EPs in morphologically salient position (yet with no morphological content), are realized phonologically as [o], which is not the usual default quality of Italian EPs (e.g. ho chiamat- + [o], 'I called'). As Repetti (2012) reports, the usual default EP is [i] in non-final position, and [e], [ə] or [ᵊ] in word-final position (although, in other variants of Italian, such as Old [pre-modern] Italian and American Italian, [o] and [a] are also attested).

Furthermore, some linguistic constraints, that regulate the production of vowel insertion in Italian have been established in studies conducted by Broniś (2016) and Grice et al. (2015). Stress, voicing of the consonant, syllable ratio, and intonational contour are all constraints that have been found to have an effect in the production of vowel insertion in Italian. Moreover, Grice et al. (2015) affirm that there is high speaker-dependent variation in the adaptation of loanwords in Italian. However, its conditions are still something unexplored. Previous studies on other languages have already established that the first and second language of a speaker interact (Baker 2005; Flege 2007; Flege et al 2003) and that knowledge of the L2 has some effect on how loanwords are adapted (Kang & Schertz 2017; Kwon 2017; Nomura & Ishikawa 2018). Moreover, stochastic models of speech processing theorize that speech categories are formed and modified thanks to repeated input (see §2.1). The main aim of this research is then to establish whether the production of vowel insertion in Italian is influenced by the amount of native English input received through both passive and active interaction. A secondary aim is to confirm the influence of linguistic constraints such as voicing and type of consonant in the production of vowel insertion, following Grice et al. (2015).

For the purpose of the research, Italian offers an ideal environment, since native English input can be controlled and estimated to a much greater extent, with respect to the rest of Europe, for instance. The Italian audience does not generally come in contact with native English speakers, as English in school is usually taught by Italian speakers, and foreign television and cinema are always dubbed. Therefore, native English must be purposely sought. Younger generations do so by spending time abroad, taking university courses taught by native English speakers, or watching movies and TV shows in original language on the internet. This makes estimating the input they received possible, as people are well aware of when they are listening to something in English because they purposefully looked for it. The conscious effort that speakers make to obtain L2 input ensures a certain degree of motivation and engagement, which are essential in language acquisition and language change (Rice 1983; Stuart-Smith et al. 2013).

2. BACKGROUND

2.1 Stochastic models

Establishing how language is processed and acquired are two of the core questions of linguistics. In phonology, these questions are translated into determining how phonological categories and processes are formed and applied. One of the most influential frameworks that try to model this is certainly Optimality Theory (OT), introduced by Prince & Smolensky (1993 [2004]). OT builds on the hypothesis that each individual possesses the same set of universal constraints, which are ranked differently depending on the language. This implies that certain constraints are in conflict with each other and are violated. In fact, the violation of constraints is a key element of OT. Violation of lower-ranked constraints is preferred over violation of higher-ranked ones, therefore, higher-ranked constraints defeat lower-ranked constraints so as to obtain the attested output in the target language (Kager 1999; McCarthy 2008).

Within OT, various approaches have been trying to explain language acquisition. One proposes that the learning of grammar is explained by (Recursive) Constraint Demotion (Tesar 1997; Tesar & Smolensky 1998), which is an algorithm that ranks constraints based on the grammar dataset, which consists of winning/loser pairs. At each pair, all constraints that disfavor the winning candidate are demoted with respect to constraints that favor the winning candidate or are neutral. Although Recursive Constraints Demotion is claimed to generally arrive at a ranking that is compatible with the grammar, it fails to account for the variation and sometimes imprecision of natural data, for it requires a dataset that is always coherent with the grammar rules and lacks variation (Tesar 1997; Albright & Hayes 2011).[1] Thus, it fails to learn realistic data. Other criticisms made towards this algorithm, although not particularly relevant to the scope of this research, are the fact that the set of constraints is given in advance, and that whenever two or more sets of grammar are found compatible, there is no device to decide which one is better (Albright & Hayes 2011).

An OT approach that does account for variation (and gradience) is stochastic OT, which employs the Gradual Learning Algorithm (henceforth GLA) (Boersma 1997). In this framework, each constraint is given a numerical value, and the "disharmony of a constraint at evaluation time is randomly distributed about this value" (Boersma 1997: 43). It follows that variation is possible in virtue of the finite numeric difference between two constraints: the smaller the difference, the

[1] One of the reasons the (Error-driven) Constraint Demotion fails to converge is that the algorithm requires the grammar to be fully-ranked, that is, constraints cannot have the same ranking in the hierarchy (Tesar & Smolensky 1998).

greater the chance of variation (for an explanation, see Boersma & Hayes 2001). In learning as well, the GLA ensures that children learn the grammar with the same amount of variability as their parents. It also works differently than Recursive Constraint Demotion, as the GLA does not rerank the constraints at every input/output pair, but causes only small changes in the value of the constraints involved (Boersma & Hayes 2001). Moreover, it both demotes the constraints that disfavor the winning candidate, and promotes those who favor it (Boersma & Hayes 2001; Coetzee & Pater 2011). However, this algorithm is not perfect either, failing to find compatible grammars with certain kinds of constraint violations (Albright & Hayes 2011).

Another theoretical framework relevant to this study is Neural Networks (NNs), as modeled by Boersma et al. (2013). As opposed to stochastic OT, NN do not present constraints or rankings, but nodes and weights (the notion of weights was already present in Hayes & Wilson's [2008] Maximum Entropy Model), aimed at resembling the brain's neurons and synapses. For this reason, NNs are one of the most biologically plausible frameworks. Let us take phonological categorization as an example to make how the NNs work clearer. Figure 1 from Boersma et al. (2013: 22) represents an NN model before any input has been received. As can be noted, all possible sounds at the auditory form ([AudF]) are placed in a continuum and are connected to all possible phonemic categories at the surface form (/SF/). All connections have the same thickness, which represents their connection weight. The connection weight can be biologically seen as the firing rate of a neuron: the stronger the connection, the higher the firing rate (Boersma et al. 2013), which results in a higher association between the two nodes.

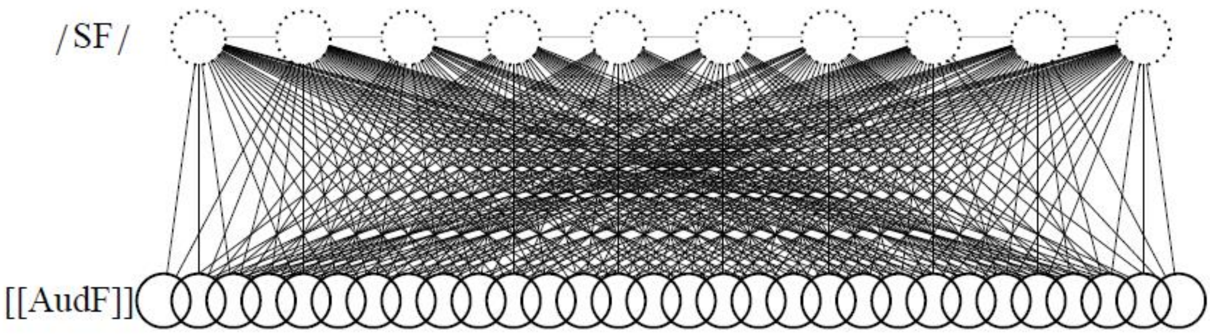

Figure 1. Network before being given any input. Taken from Boersma et al. (2013: 22)

Categorization is achieved through presenting the model with a great number of auditory values at the auditory form. With the repetition of input, particular values in the AudF continuum are associated with categories at SF, and the connections with those categories become proportionately stronger. Variation is possible depending on the strength of the connections. If for example, with a particular AudF value *x*, a connection to a category A has strength 0.10, the connection to category B has strength 0.00, and the connection to category C has strength 0.90, the learner will identify the AudF value as A 10% of the time, as B 0% of the time, and as C 90% of the time (see Figure 2).

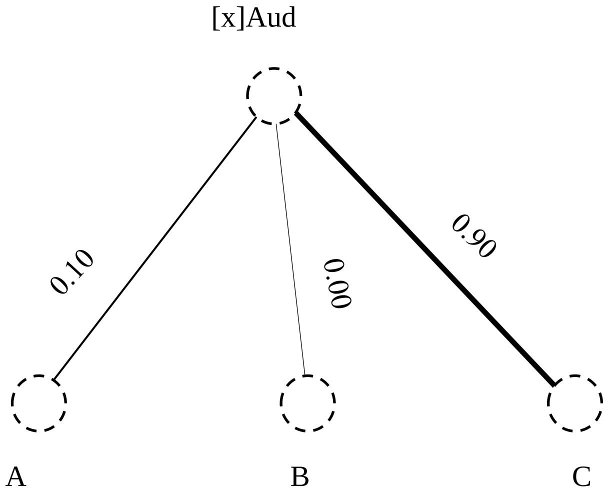

Figure 2. Gradience in category perception in NN.

The fundamental point of NNs, and other theoretical proposals such as stochastic OT, Maximum Entropy Model (Hayes & Wilson 2008), and other stochastic-based ones, is that learning is achieved through the statistical parsing of speech acoustics, that is, through the recognition of patterns thanks to repeated input. This poses the basis for predicting that categories or constraints can be modified or formed if enough input is received. This pivotal concept is being applied in the present research as well. Can Italian adult speakers change the way they perceive and produce releases in English loanwords, provided they received enough input from the source language? This question will be empirically answered, by estimating the amount of input received by the participants of the study and correlating it with the amount of paragogic vowels produced.

The next sections will discuss previous literature regarding the interaction between phonetic subsystems of different languages, the difference between canonical EPs and intrusive vowels, and Italian vowel insertion. Moreover, the reliability of one of the measurements used in the methodology, language self-assessment, will be briefly discussed as well, due to its controversial status within the literature.

2.2 L1-L2 interaction in loanword adaptation

It is an established fact that the second language (L2) of a speaker is deeply influenced by their first language (L1), above all in the early stages (Hancin-Bhatt & Bhatt 1997). It has also been established that the L2 can have effects on the L1 at all levels, for example in language attrition (see Köpke et al. 2007 [eds.]). In fact, the two linguistic systems interact with each other and can display some overlap. Regarding the phonetic domain, Flege et al. (2003) found that phonetic category assimilation/dissimilation in early and late bilingual Italian-English speakers varied based on their age of arrival in the English-speaking country (Canada) and their frequency of use of their L1 (Italian). In English production, late bilinguals would assimilate the English vowel [eɪ] to the Italian [e], thus producing [eɪ] with less formant movement than native English speakers. On the other hand, early bilinguals with low use of Italian would produce [eɪ] with too much formant movement, thus dissimilating the two sounds. Thus, while late bilinguals do not seem to be able to keep the two systems separated, early bilinguals are more likely to do so. Similar findings were reported in other studies, such as Baker (2005) and Flege (2007), among others. In particular, the latter reaches the conclusion that age of arrival, hence the plasticity of the brain, is not the only factor that contributes to a native-likeness of the L2 phonetic system, but also L1 interference and, crucially, frequency of use (which correlates to the amount of input received).

The studies mentioned above all dealt with immigrants acquiring the L2 and focused more on their performance in the L2. Our study shifts focus towards L1 performance, and the influence that experience in the L2 can have in the adaptation of loanwords from the L2 into the L1. A number of studies investigated similar contexts. Kwon (2017) found that Korean speakers with different levels of experience with the L2 (late bilinguals, early bilinguals and monolinguals) had different epenthesis patterns when adapting English non-words into Korean. English coda consonant releases usually trigger the insertion of an epenthetic [ɨ] in the Korean adaptation. In that study, the insertion of the vowel depended in part on the level of experience with the L2. For example, monolingual and late bilingual speakers, hence with less experience in English, would insert the vowel more often than early bilinguals. Another study by Kang & Schertz (2017) points towards loanword adaptation not merely being the application of L1 phonotactics to loans, but also being influenced by the speakers' knowledge of the L2 phonology. They found that Korean speakers were more consistent in mapping English voiceless plosives onto Korean aspirated plosives (as opposed to tense or non aspirated) when their L2 proficiency was higher. The latter was calculated based on the similarity of their perception of L2 categories to the native speakers' one. They concluded that as one's knowledge of the L2 advances, L2 speakers rely more and more on L2 categories in cross-language mapping. Finally, Nomura & Ishikawa (2018) reached similar results in a perception study with Japanese learners of English. Their proficiency level was determined through a standardized test that they took independently prior to the study, and they were divided into the groups 'introductory' and 'intermediate'. They demonstrated that intermediate speakers perceived an epenthetic vowel in English loans less frequently than introductory speakers, meaning that the more advanced speakers had a more native-like perception.

2.3 On the status of Italian vowel insertion

An important terminological distinction needs to be established, before going any further. EPs may or may not have a phonological role, and native speakers may or may not perceive them (Hall 2011). Hall (2006) distinguishes between canonical EPs, which are perceivable by the native speakers of the borrowing language, participate in phonological processes (i.e. they repair a violated structure), and have articulatory characteristics similar to lexical vowels. On the other hand, 'excrescent' or 'intrusive' vowels are invisible to phonological patterns, the vowel is usually a schwa or a copy of a neighboring vowel, it usually occurs in heterorganic clusters, and it is likely optional.

Several of the studies dealing with EPs in Italian are actually referring to EPs perceivable by native speakers (Cardinaletti & Repetti 2007; Repetti 1993, 2006; see Passino 2008 for an account on EPs and consonant gemination in word-final position; see Repetti 2009 for an account on EPs and consonant gemination in American Italian). As mentioned before, word-final epenthesis is reportedly [e], [ə] or [ᵊ]. Repetti (2012) investigated the quality of these EPs compared to the quality of the lexical vowel /e/. Her findings suggested that (i) these EPs are optional, (ii) when present they do not contribute towards rhyming with e-final nouns (for example, 'tram' [tramme/trammə/trammᵊ] does not rhyme with a word like 'fiamme' [fjamme]), (iii) their first and second formants do not correspond to those found for any Italian lexical vowel, being more centralized like a schwa. (ii) and (iii) clearly show that these EPs do not have the same phonetic properties as lexical vowels. Moreover, she reports that (iv) the speakers did not perceive them either in perception or production, (v) they do not have a (morpho)phonological role or take part in phonological or morphological processes. She concludes that the inserted vowels are not the result of a violated structure that needs repairing but that they are "part of the release of the word-final consonant" (Repetti 2012: 175), meaning that Italian consonants need a strong release to be perceived as such, and this is sometimes achieved in word-final position with vowel insertion.

Although some authors regard these inserted vowels as phonological, mainly due to consonant gemination in final position (Bafile 2005; Broniś 2016; Passino 2008), I am going to follow Repetti's (2012) conclusions, due to the optionality and non-perceivability of the inserted vowels.[2] Thus, this type of vowel insertion is here categorized as the one described by Hall as the 'excrescent' or 'intrusive' type of EP. For the sake of clarity, from now on we will refer to phonologically active epenthetic vowels as EPs, while non-perceivable EPs will be called 'intrusive vowels' (henceforth IVs), following Hall (2006, 2011). Please note that in the literature, oftentimes no distinction between these two types of epenthesis are made.

2.4 Linguistic constraints on Italian intrusive vowels

As mentioned in §1.1, there has been a number of studies tackling Italian EPs (Cardinaletti & Repetti 2007; Repetti 2006, 2009). In this section, we are going to address studies dealing with IVs, and in particular with the linguistic constraints that regulate the production of IVs. Aside from the study conducted by Repetti (2012), which we have already described in §2.3 in order to make a distinction between EPs and IVs, two other studies are worth describing. The first one is by Grice et al. (2015). They conducted an experiment with 10 participants speaking the Bari variant of Italian (southern macro-area), aimed at investigating the effects of prosody on the production of IVs. Their

[2] All the studies mentioned above focus on varieties of Italian that differ with each other and with this study. In fact, the majority of these studies base their claim on data from central or southern varieties of Italian. The present study, however, is based on data from the Veneto region. Hence, what is claimed to be an IV in this study could be an EP in another, due to different phonotactics. This being said, determining whether vowel insertion in Italian loanword adaptation is a phonological or phonetic process is beyond the scope of this paper.

results showed that the presence of IVs is favored by a raising intonational contour (as opposed to falling), voiced consonants (vs voiceless), and smaller syllable ratio per word (monosyllables had more IVs than bisyllables). Another study is the one conducted by Broniś (2016) on 38 speakers of the Roman variant of Italian (central macro-area). Her dataset suggests that IVs are inserted after loanwords ending in clusters, and are favored when stress falls on the stem-final syllable. However, it must be noted that the statistics employed is suboptimal, therefore the results must be taken cautiously.

Studies and phonological textbooks, such as Nespor (1993), and Rabanus (2003), state that only sonorants, as opposed to obstruents, are allowed in coda position in Italian. In fact, the only words that end with a consonant in Italian, that are not loans, end with a sonorant ('buon' [good], 'per' [for], 'non' [not], 'Dottor' [Doctor] etc.). This could lead to the prediction that vowels should be inserted to a smaller extent after sonorants, especially if we consider them to be the canonical EPs that take part of phonological processes, as some authors claim (see §2.3). In this view, sonorants would not violate any constraints when in coda position, hence would not need the insertion of an EP. Crucially, Grice et al. (2015), with a random forest analysis, found that type of consonant (sonorant vs obstruent, and stop vs fricative) was the factor that impacted the production of IVs the least, among all the factors taken into consideration, and concluded that there appeared to be no independent effect for type of consonant.

Nevertheless, it must be considered that the variant tackled by that study is a southern Italian one, with different constraints than the variant used in the present study, a northern one. It is known that regional variants of Italian are deeply influenced by the local minority language spoken there (Berruto 1989), and Venetian, the minority language spoken in Veneto, regularly allows sonorants in word-final position (see Ursini 2011). Therefore, there might an effect on the production of vowel insertion that was not detected in Grice et al.'s (2015) study, due to the variant of Italian utilized. For this reason, type of consonant will be included in this study as a predictor.

2.5 Second language self-assessment proficiency

Language self-assessment is widely used as a tool to measure language proficiency in linguistic studies. However, previous literature has been contradictory regarding the validity and reliability of self-assessment measurements of proficiency. While many studies have found positive correlations between self-assessment and proficiency tests (Ashton 2014; Brown et al. 2014; Edele et al. 2015; LeBlanc & Painchaud 1985; Mistar 2011, among others), there is a divided opinion on whether this tool is actually reliable.

The reliability of L2 self-assessment is challenged by the fact that the majority of sources agree that self-assessment shows some bias, for example participants tended to over- or underestimate their L2 competency based on gender or cognitive ability (Edele et al. 2015). It has also been demonstrated that proficiency itself has some effect on how reliable self-assessment is (Ashton 2014; Brown et al. 2014; Ross 1998). In particular, they found that low-proficiency speakers are not able to assess their own proficiency levels as accurately as higher-proficiency speakers.

The accuracy of L2 self-assessment is considerably improved when more specific and elaborate types of test are employed (Edele et al. 2015). Tests based on can-do statements were found particularly reliable (Brown et al. 2014). Can-do statements are descriptions of concrete tasks and situations that the learner is likely to have encountered. They are usually ordered based on difficulty, with tasks such as 'writing a postcard' listed in the beginner levels, and 'understanding a movie without subtitles' listed in the advanced levels. The learner is asked to indicate their ability in doing each task. Within the Common European Framework of Reference (CEFR), six proficiency levels are distinguished, and described with concrete tasks that the learner should be able to do at each proficiency level (Council of Europe 2011).

3. PRESENT STUDY

So far, the studies presented showed that the adaptation of loanwords is filtered through the L1, but can be influenced by the L2 as well. The extent of L2 influence depends on the experience of the speaker with the L2. Experimental studies have measured L2 experience in different ways. Kwon (2017) correlated experience with bilingualism, Flege et al. (2003) and Flege (2007) with age of arrival and frequency of use, while Kang & Schertz (2017) and Nomura & Ishikawa (2018) used L2 proficiency, albeit measured with different methods. On the other hand, theoretical frameworks such as the ones previously presented, and especially NNs (Boersma et al. 2013), predict that it is the amount of input that directly shapes categories and constraints, and other measurements are merely correlated. For obvious reasons, it is not feasible to calculate the amount of input speakers receive in the course of their life. However, the conditions in Italy are such that it is possible to make an estimation of English input received in a certain window of time. Therefore, the present study aims at establishing whether the production of IVs is controlled by the amount of input received in English. The novelty of this research lies not only on this method, but also on the contraposition made with L2 proficiency. A second aim of this study is to verify findings from other studies regarding the influence of voicing and type of consonant (obstruent vs sonorant) in the production of IVs.

4. METHODS

4.1 Participants

The experiment featured 21 native Italian speakers, recruited through word of mouth, through university internal e-mails at the University of Padua or posts in Facebook groups of the University of Padua and Ca’ Foscari University of Venice. They are all university students (4 males, 17 females) aged between 19-39 yo (mean = 25;0 yo), born and raised in Veneto, with Northern Italian parents. They have different levels of proficiency in English, ranging from beginner to advanced, although the majority has intermediate competence in the language.

4.2 Materials

The experiment was divided in two parts. The first was an irregular plural elicitation task (following Repetti 2012), in the form of a PowerPoint presentation (Appendix A). The participants were showed a countable Italian word and had to produce the plural, as in examples (1) and (2). They were asked to pronounce both the singular, as displayed in the screen, and the plural, preceded by the numeral. The task took advantage of the fact that consonant-ending English loanwords do not show plural inflection in Italian. In short, the singular and the plural form for these words are identical, allowing to collect double the data.

(1) Una casa → due case

‘One house → two houses’

(2) Un tunnel → due tunnel

‘A tunnel → two tunnels’

The task was made up of different types of words, ordered randomly: 75 of them were English loanwords commonly used in Italian, while 50 of them were fillers, consisting of Italian nouns whose plural can be either regular or irregular. The loanwords were selected from the judgements of the author, a native Italian informant, and an Italian dictionary, in an attempt to control for the nativization of the words, i.e. whether they would be perceived as naturally belonging to Italian or as unfamiliar loans. First, a pool of words that were judged as commonly used in the Italian language by the researcher was formed, and then checked by a native Italian informant. Consequently, the existence of each word was checked in *Il Nuovo De Mauro* dictionary (De Mauro 2018), which reports the period in which the loanword has first been used in an Italian context. About two thirds of the words chosen have been incorporated in the Italian language for 50 years or more. The majority of words that have been incorporated later belong to the technology semantic field, and include extremely frequently used words such as ‘computer’, ‘chat’, ‘web’, ‘password’, ‘blog’ and ‘mouse’. Other words, such as ‘skateboard’, ‘piercing’, come from the 80s and 90s pop culture, and are still frequently used. In any case, all the loanwords were incorporated in the Italian language at least 20 years ago. Words that did not exist in the dictionary, such as ‘location’, or ‘muffin’, were excluded. Other words that did exist in the dictionary, such as ‘chairman’ or ‘challenge’, but were not judged frequently used by the author, were also excluded. Finally, three words that were not included in the dictionary, but were judged extremely used by the younger generations were included. These words are ‘tablet’, ‘hashtag’ and ‘iPad’.

The loanwords selected have different numbers of syllables (between one and three), and are all countable, so that producing the plural sounds more natural. They are all consonant-ending words, and are subdivided in three groups, depending on the quality of the final consonant: voiceless plosive or fricative, voiced plosive or fricative, and sonorant consonant. The total amount of data collected is therefore 3150 tokens, while the number of tokens analyzed is 2900.

The second part of the experiment was a sociolinguistic guided questionnaire (Appendix B), which inquired on general sociolinguistic information (i.e. age, languages spoken, discipline studied at university etc.), and tried to estimate the amount of native English input received through both active and passive interaction in the previous years. The questions are directed towards estimating an amount in hours as precisely as possible of native English input received through various sources, including periods spent in English-speaking countries, the vision of movies and TV shows in the original language (with and without English subtitles), and other regular interactions with native English speakers.[3]

[3] We decided that vision of media with Italian subtitles would not be included in the count. In fact, while it is attested that intralingual subtitling increases one’s proficiency (Borrás & Lafayette 1994), it is not yet clear whether subtitles in

The participants were also asked to assess their own English proficiency (Appendix C). Testing English proficiency with an English test was avoided, due to time restrictions, and the fact that the entire experiment was in Italian and on the Italian language. Based on previous literature, which found that can-do statements are reliably correlated to language proficiency (see §2.5), we decided to adopt the CEFR indicators as a way to measure English proficiency (Council of Europe 2011). In the can-do self-assessment task, the participants had to choose one of six indicators corresponding to the levels of the CEFR (from A1 to C2) for each of the five fields: reading, listening, oral interaction, writing and speaking. Speakers were also allowed to choose two adjacent levels in case they were doubting which level was the most suited. Each level was then assigned a grade from 1 (for A1) to 6 for (C2) and the average proficiency for each speaker was calculated. Speakers that scored less than 4 (B2) fell in the 'low proficiency' group, while those who scored 4 or above fell in the 'high proficiency group'[4].

4.3 Procedure

The experiment was carried out at the phonetic lab of the ISTC (Istituto di Scienze e Tecnologie della Cognizione), part of the CNR (Consiglio Nazionale delle Ricerche) in Padua, Italy. The participants first signed a consent form, which allowed for the experiment to be recorded. The first part of the experiment was the plural elicitation task presented in a slideshow in which every slide shows a single word. The participants were asked to produce the plural of the word shown on the screen after having read it aloud, as mentioned in the previous section. The participant could then click a button in the keyboard to proceed to the next slide. After the elicitation task, the participant

another language would have the same effect. Moreover, it has been attested that watching media content subtitled in the mother tongue is mainly useful only in the first stages of L2 learning (Almeida & Costa 2014).

[4] It was decided to divide the speakers in two large proficiency groups, mainly due to the method employed to measure proficiency. Although the self-assessment proficiency test here employed was considered reliable enough to distinguish between speakers with high or low proficiency in English, it was not considered reliable enough to divide speakers in such precise groups as the CEFR levels. However, this less precise division in two groups still remains less reliable for speakers with scores close to the cut-off point than speakers that scored further from the cut-off point. See also note 6.

was asked to answer the questions present in the sociolinguistic questionnaire. The format was that of an interview: the experimenter would ask some questions following the questionnaire and the participant would answer. The answers were written down by the experimenter herself. This method allowed for both the experimenter and the participant to ask for clarifications and overall for a more precise estimate to be calculated. Finally, the participant would fill in the English self-assessment task.

4.4 Analysis

Data analysis was carried out with the program *Praat* (Boersma & Weenink 2018), in which IVs were identified and segmented. Establishing criteria for the presence of IVs was made necessary not only due to the fact that they are not easily perceivable by non-trained native Italian speakers, but also because vowels are in part defined language-specifically. Speakers of different languages have in fact different criteria for the perception of vowels, and phonemes in general. For example, in the adaptation of English loanwords into Korean, the release of a consonant is enough phonetic material for Korean speakers to identify it as a vowel, due to the fact that Korean plosives are not released (see Boersma & Hamann 2009a for a detailed account on loanword adaptation in Korean). Likewise, Italian IVs are identified as vowels by English speakers, but not by Italian speakers. The presence of vocalic elements was therefore determined through the following general criteria, which are applicable to both languages here concerned (Italian and English):[5]

- Presence of a periodic waveform
- Presence of voicing
- Presence (and movement) of formants
- Intensity

[5] Please note that these criteria are not enough to distinguish between phonetic and lexical vowels in Italian. These criteria are also applicable to sonorants. In case the consonant in word-final position was a sonorant, the presence of an IV was determined through formant movement.

- Duration longer than 10 ms.

A vocalic element in word-final position was signaled by a ‘YES’ point in the first tier of a text grid, while its lack was signaled by a ‘NO’ point. A second tier was used to indicate and segment the whole word, while a third tier was used to indicate whether the intonational contour was rising or falling. Finally, a fourth tier was used to segment the duration of the vocalic element, when present. A *Praat* script (given in Appendix D) then extracted all this information and created a table.

4.4.1 The problem of /r/

Rhotics are generally particularly hard to segment, compared to other type of consonants. One of the reasons is that elements with the feature [+ rhotics] are not cross-linguistically associated by any phonetic parameter, or articulatory correlate (Lindau 1980, Hayward 2013). It is also well-known that rhotics vary greatly even within the same language (Baltazani & Nicolaidis 2013). The Italian /r/ has two allophones, the flap [ɾ] and the trill [r], which are distinguished by the number of taps on the palate. Moreover, the region of Veneto, where the data was collected, presents a regional variant of the phoneme /r/, realized as a postalveolar, retroflex flap, which tends to be lateralized (Romano 2013: 218). The realization of rhotics can also in part depend on the amount of airflow employed. Catford (2001) reports that a low amount of airflow may result in the rhotic being realized as a fricative, and an even lower amount could result in it being realized as an approximant.

Indeed, a great amount of variation was also found in this corpus, for all the reasons noted above. It was necessary then, to come up with criteria for the recognition and segmentation of the vocalic elements that could follow a rhotic. In the fricative and approximant realizations of the rhotic, the criteria given in §4.4 apply. For the trill [r] and flaps [ɾ] and [ɻ], it was decided to segment the vocalic element right after the last visible tap.

4.4.2 *Criteria for token exclusion*

Despite all the precautions taken in selecting the loanwords used in the task, occasionally a speaker would produce a plural following English morphology or using an evident English pronunciation, for example pronouncing the aspirate /h/ when the grapheme ‘h’ is present, or employing the postalveolar approximant [ɹ], instead of the Italian flap [ɾ] or trill [r]. These were taken as evidence that the speaker did not recognize the word as Italian and code-switched to English, probably due to their (un)familiarity with the word and the frequency of usage. As this was judged to be speaker-dependent, i.e. no word was consistently judged unfamiliar by a high number of speakers, the tokens which were pluralized following English morphology or pronounced with English phonetics were excluded from the study. This is the primary reason tokens were excluded.

Other tokens were excluded due to:

a. connection of the consonant-ending loanword to a subsequent word beginning with a vowel, as in example (3) and Figure 3. In such cases, the insertion of a vocalic element would be unnecessary, as the following vowel acts as an effective substitute.

(3) …Due hostess_una risata…

‘…two hostesses_one laughter…’

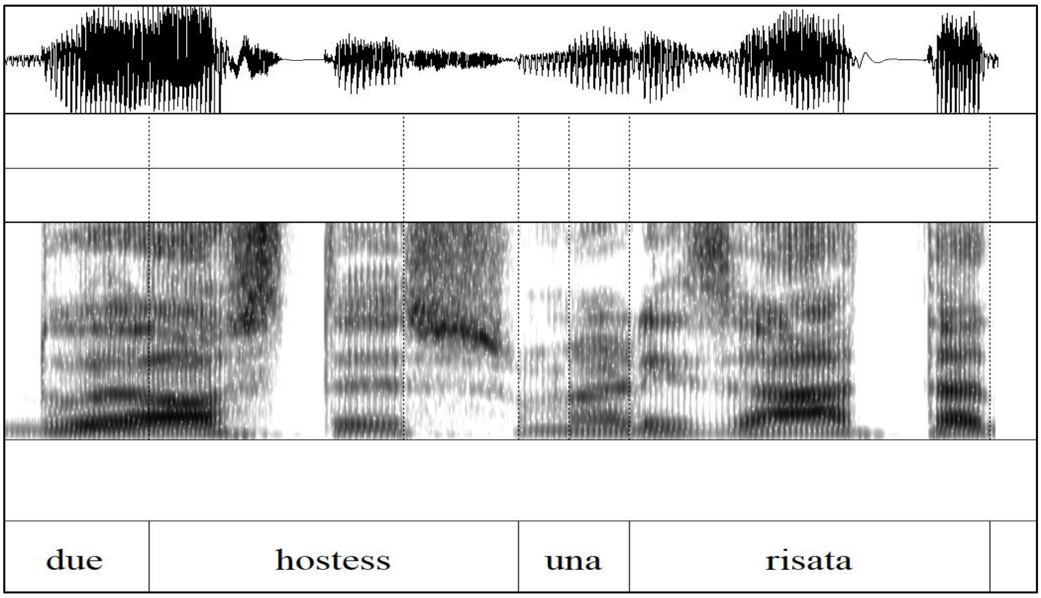

Figure 3. Spectrogram and text grid of example (2). (Speaker AA_06)

b. Connection of the loanword to a following identical phoneme, as in example (4), Figure 4. This resulted in the target token being geminated and merged into the following word.

(4) …Un tabloid_due tabloid…

‘…One tabloid_two tabloids…’

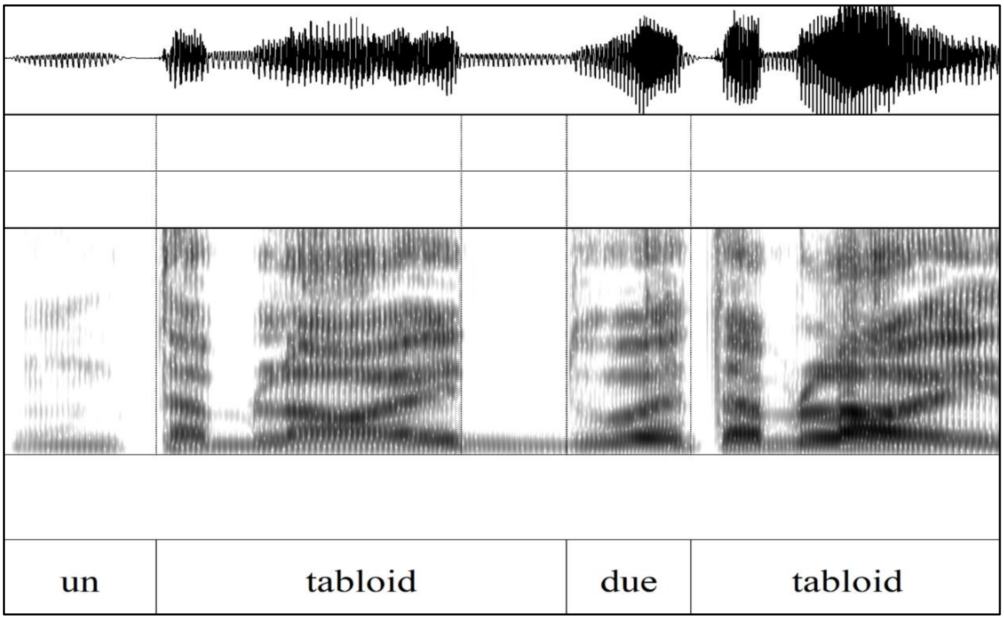

Figure 4. Spectrogram and text grid of example (3). (Speaker AA_06)

c. Nasalization of the vowel. In loanwords ending with the sonorant /n/, the vowel would merge with the nasal, resulting in a nasalized vowel, as Figure 5, depicting the loanword ‘fan’. This happened in a very limited number of cases.

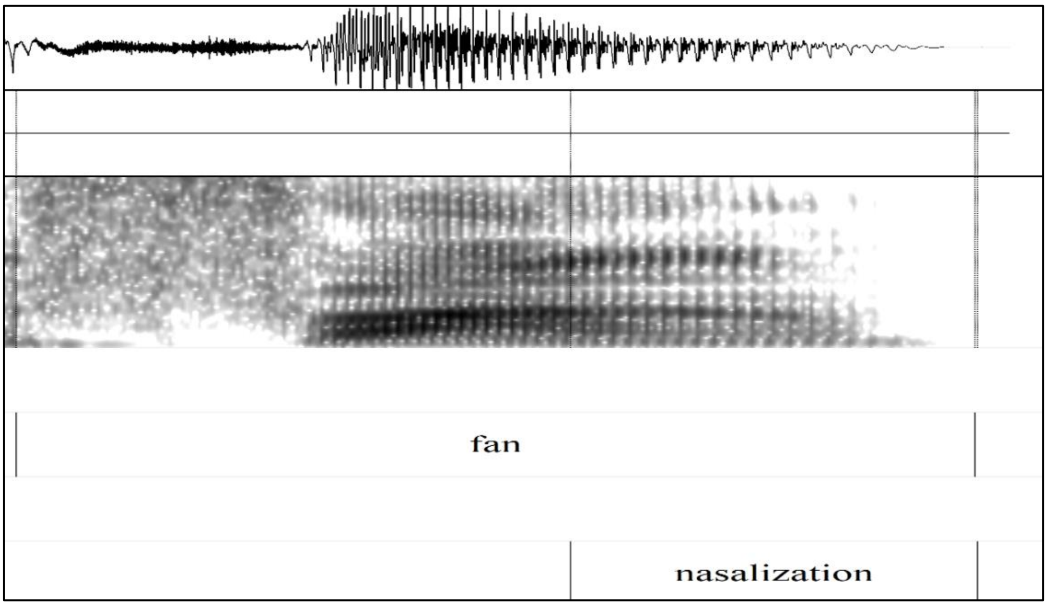

Figure 5. Nasalization of the vowel. (Speaker AA_06)

d. Deletion. At times the speaker would not realize the target consonant, as in Figure 6. This happened mostly in case of word-ending clusters, such as ‘trend’ or ‘boomerang’, but instances of deletion in simple codas were also found, for example with the word ‘canyon’. Moreover, in a couple of instances the word-final plosive was not released. This also counted as deletion, as all consonants in Italian have to be released to be perceived as such.

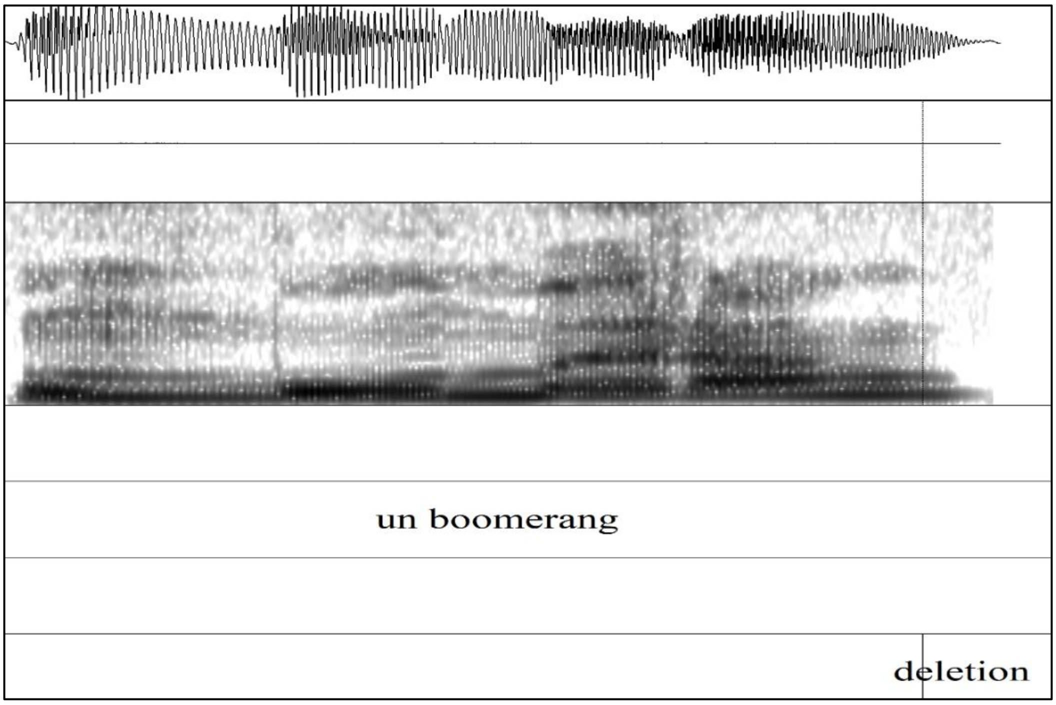

Figure 6. Deletion of the target consonant. (Speaker BI_11)

e. Mispronunciation, filler sounds or laughter. These phenomena as well caused the exclusion of some tokens. In particular, filler sounds were used when the participant was uncertain of how to produce the plural. Italian filler sounds are realized as a long schwa /ə/, /e/ (or /m/), thus are very similar to the realization of paragogic vocalic elements. Therefore, all tokens that showed much hesitation were excluded, because the vocalic element that followed was judged as being a filler.

4.5 Statistical analysis

The data was statistically analyzed with the program R (R Core Team 2016), specifically using a generalized linear mixed-effects model (glmer). The dependent variable is the **presence of the IV** (expressed through 1 for YES and 0 for NO)[6], while the predictors are the level of **English proficiency** ('high' H, and 'low' L, whose contrasts were H = +0.5 and L = -0.5)[7, 8], the amount of

[6] Called 'EP' in the model.
[7] Called 'proficiency' in the model.

English input received (a continuous numeric variable, which was centered in the model)[9], and **type of consonant** ('sonorant' SN, 'voiceless obstruent' VLOB and 'voiced obstruent' VDOB, whose contrasts varied based on the analysis carried out, and are specified below)[10]. The statistical design also features 'participant', with the type of consonant as a random slope, and 'word' as random variables.

Due to the nature of the study, e.g. lack of voiceless sonorants, the data was unbalanced, forcing the model to have a three-way contrast per 'type' of consonant, i.e. between the categories SN, VDOB and VLOB, instead of two two-level contrasts ('sonorant' vs 'obstruent', and 'voiceless' vs 'voiced'). Moreover, it was necessary to drop 'tone', e.g. a rising or falling intonational contour, as a factor and random slope, due to the model failing to converge. This was probably caused by a very skewed distribution of the intonational contour, with many participants adopting a rising tone for both words within the same pair, instead of a falling-raising or raising-falling pattern. The model in (5) is therefore the model used in this study (see the R coding in Appendix E).

```
(5)  model <- glmer (EP ~ inputC * proficiency * type + (type | speaker) + (1
   | word), family=binomial, data=table, control = lmerControl (optCtrl = lis
   t (maxfun = 100000))
```

[8] As mentioned before, it was decided to divide the speakers in two large proficiency groups, instead of using proficiency as a continuous predictor. Binning the data essentially means that information is lost in the statistical analysis. According to Wainer et al. (2006), data binning could lead to show trends which are not reflected in the underlying data. However, this problem can be solved if we look at the scatterplots of the unbinned data, which are provided in §5.1.

[9] 'InputC' in the model (centered).

[10] 'Type' in the model.

5. RESULTS & DISCUSSION

5.1 Input vs proficiency

Figure 7 shows the overall distribution of speakers, according to the percentage of the IVs produced (abscissa) and the amount of English input received (ordinate). As can be seen, the scatterplot indicates that the speakers' tendency to decrease the percentage of produced IVs is inversely proportional to the increase in the native English input received.

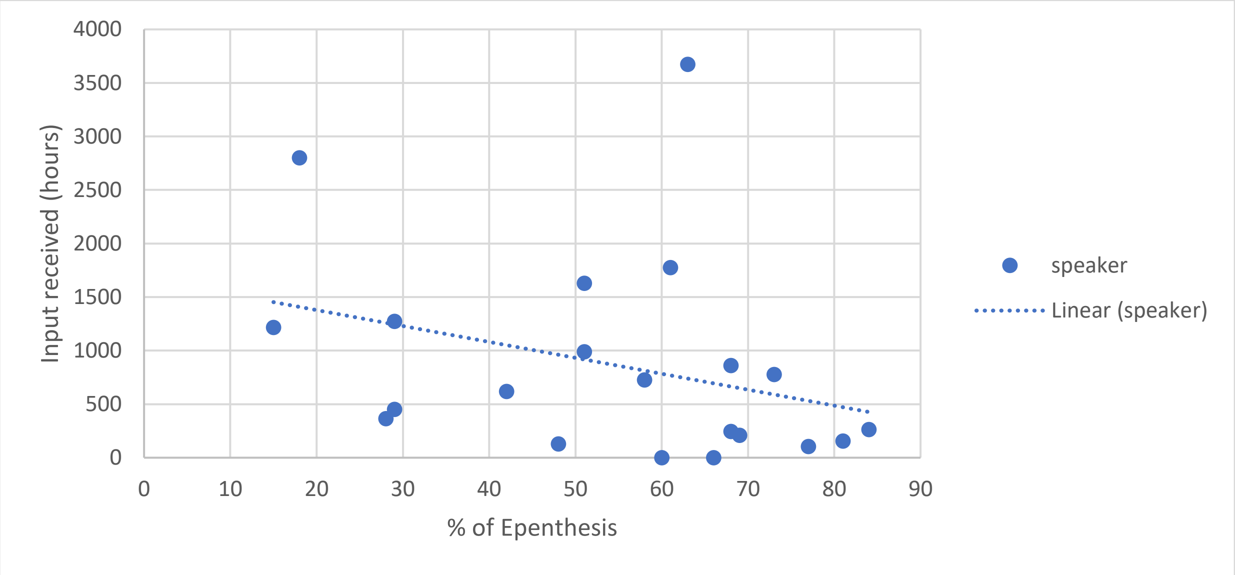

Figure 7. Scatterplot of speakers per input received and percentage of IVs produced.

In Figure 8, speakers are distributed according to the percentage of IVs produced, and according to their levels of proficiency. The level of proficiency here shown is the arithmetical average of the answers they gave in the self-assessment proficiency test, on a scale from 1 to 6. Please note that in the statistical analysis the speakers were simply divided into the groups 'high proficiency' and 'low proficiency'. Numerical scores are used in this graph for comparison purposes. The scatterplot shows an inverse correlation between English level and production of IVs. However, it is somewhat weaker than the correlation shown in Figure 7. Therefore, in this population, input seems a much stronger predictor than English proficiency.

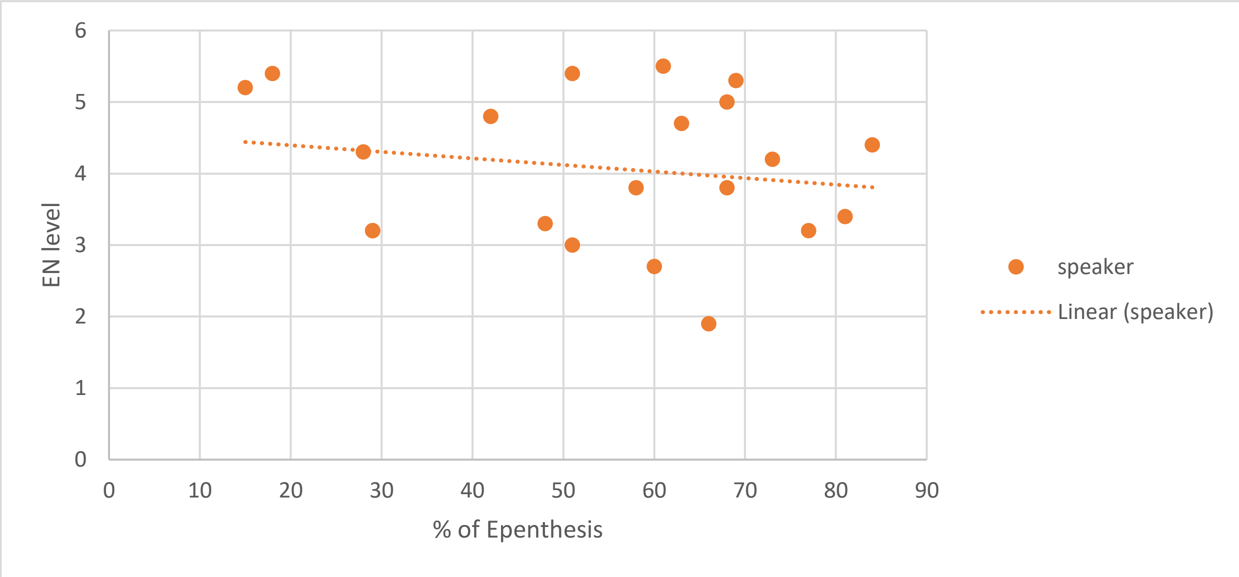

Figure 8. Distribution of speakers per level of English and percentage of IVs produced.

The statistical analysis was performed with a generalized linear mixed-effects regression. The model indicates that the odds of IVs being produced decrease by 1.1 times every 100 hours of English native **input** received (95% confidence interval [henceforth CI] 1.01..1.20 times, *Pr(>|z|)* = 0.039). This means that the native English input does contribute to decreasing the percentage of IVs produced, confirming our prediction. On the other hand, **proficiency** does not have significant results (estimate = 1.77 times, 95% CI 0.46..6.80 times, *Pr(>|z|)* = 0.39), therefore no generalizations are possible. Although the estimate points counterintuitively towards high proficiency speakers producing more IVs than low proficiency speakers, the *z*-value is not significant, and the CI too wide, with it extremes pointing in two different directions. It is also not possible to make a comparison with input, due to the proficiency CI including the input CI.

However, we can undoubtedly state that L2 input has an effect, albeit small, on IV production. In fact, it takes hundreds of hours of listening or oral interaction (i.e. likely hundreds of thousands of tokens) to produce a change in quality of production of word-final consonants, with fewer consonants requiring an IV to reinforce the release. As it is known that children take years of input to build and perfect their phonological system, it is realistic that it will take hundreds of hours to produce a small phonetic change such as IV production.

This study suggests that input is a much more direct way to investigate L2 influence on loanword adaptation than other measures used in previous studies (see §2.2), such as proficiency, bilingualism or age of arrival. The main reason is that by its own nature, input is an acoustic signal that is filtered by perception and stored in memory. On the other hand, L2 proficiency and other measures take into account language development as a whole. L2 proficiency especially (bilingualism and age of arrival are not applicable to this study, hence are not relevant) does not necessarily entail that the learners have regularly come into contact or acquired the language from native speakers. In fact, in Italy, this is usually not the case. This can result in speakers having intermediate/high levels of proficiency and poor L2 phonotactics. Hence, that without accounting for input, mismatches between percentage of epenthesis and proficiency levels would be left unexplained (see Figure 8).

This study also suggests that it is essential to take into account passive input received through media. In this study, input estimation is the addition of time spent in an English-speaking country, of a yearly estimation of hours spent watching media content (tv shows, videos, movies) with or without English subtitles, and of the 'other' category, which is oral interaction with native English speakers in the home country. As can be seen from Table 1, 11 out of 21 participants have never been for an extended period of time in an English-speaking country. Of these, 9 do not have other forms of regular oral interactions with native English speakers. Therefore, about half the speakers of this study received native English input exclusively through passive sources such as medias (if any). This means that their gradience in IVs production can only be accounted for by passive input.

Table 1. Calculation of input amount per speaker.

SPEAKER	ENGLISH PROFICIENCY LEVEL	TIME ABROAD (hours)	Hours of MEDIA WITH(OUT) EN SUBS	OTHER (hours)	TOT in hours
BB_01	L	0	0	0	0
AB_02	H	140	572	150	862
LC_03	L	0	546	180	726
GB_04	L	0	1274	0	1274
CD_05	H	0	468	150	618
AA_06	L	0	0	0	0
AM_07	H	360	416	0	776
GZ_08	H	0	208	0	208
MZ_09	H	720	910	78	1708
SS_10	L	0	988	0	988
BI_11	L	0	104	24	128
SF_12	H	450	676	90	1216
PI_13	H	1620	156	0	1776
FF_14	H	2400	400	0	2800
PL_15	L	360	90	0	450
MD_16	L	90	156	0	246
SR_17	L	0	156	0	156
DB_18	H	0	364	0	364
GF_19	H	0	262	0	262
MR_20	L	0	104	0	104
SC_21	H	2920	754	0	3674

5.1.1 *A BiPhon-NN explanation of the role of input in IV production*

I theorize that the change in gradience of IV presence produced by the amount of input is achieved through perception, although it is applied in production. In perception, the speaker receives input for word-final consonants that is produced with vowel insertion a certain percentage of times, and without vowel insertion the rest of the times. Figure 10 shows a simplified version of the perception of the word ‘jet’, using the BiPhon-NN model (Boersma 2007; Boersma & Hamann 2009a, 2009b; Boersma et al. 2013), which features three levels of representation: auditory form [AudF], surface form /SF/ and underlying form |UF|, depicted in Figure 9.

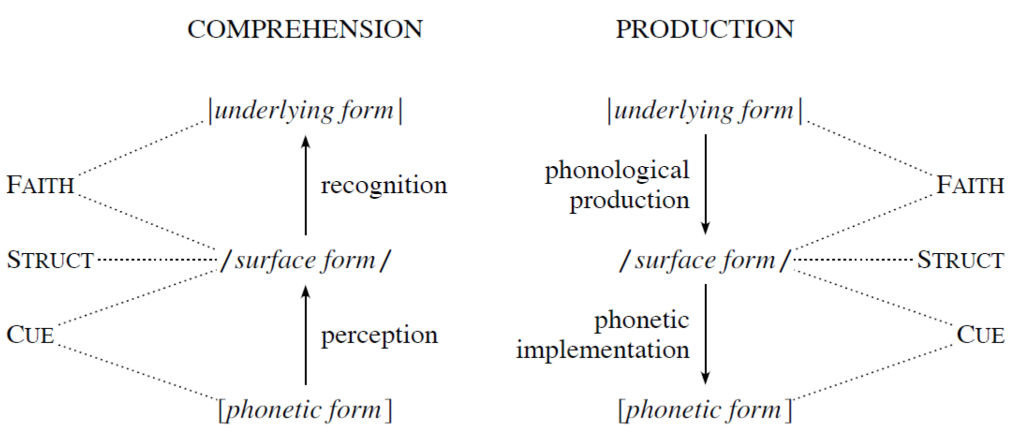

Figure 9. The BiPhon model (from Boersma & Hamann 2009a).

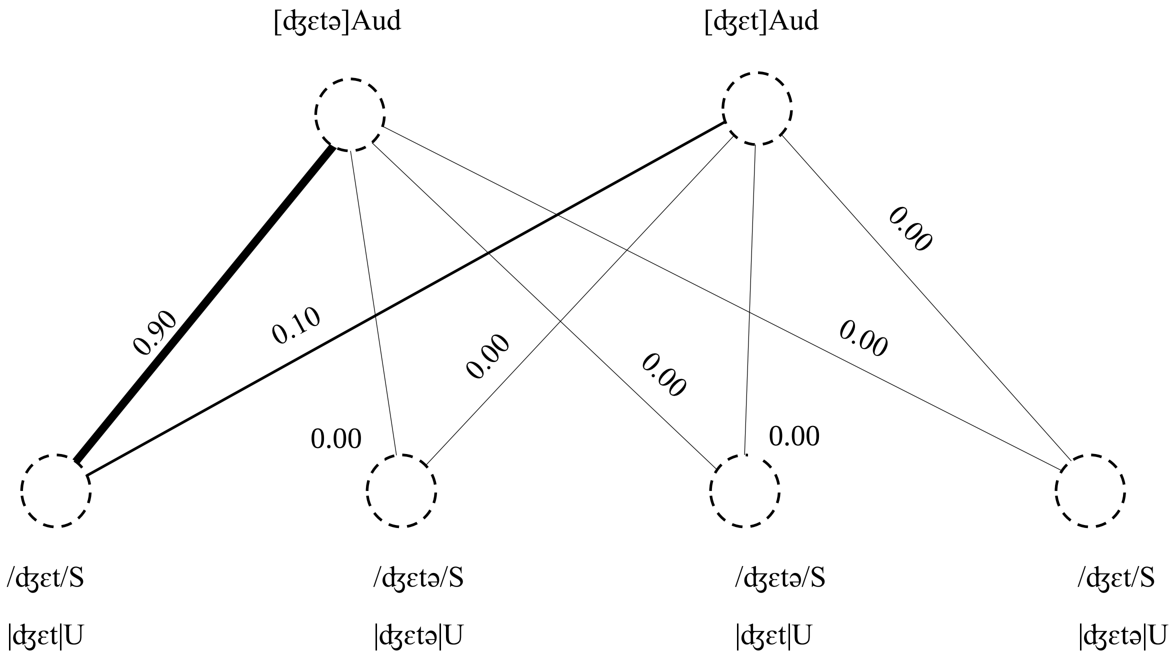

Figure 10. Speaker A's perception of the word 'jet'.

In figure 10, 90% of the input received contains an IV in word-final position, while 10% does not, resembling the hypothetical variation in perception belonging to speaker A with close-to-none contact with English. In Figure 11, speaker B received input with IV insertion 30% of the time, while 70% of the time the input received contained no IV, resembling the hypothetical variation in perception of a speaker with much contact with native English. As can be seen, in both cases the IV is not perceived because it is a phonetic vocalic element that is part of the release of the consonant (Repetti 2012)[11]. Therefore, the received phonetic signal [ʤɛtə]Aud or [ʤɛt]Aud is perceived phonologically as /ʤɛt/S in surface form and consequently stored in underlying form as |ʤɛt|U.

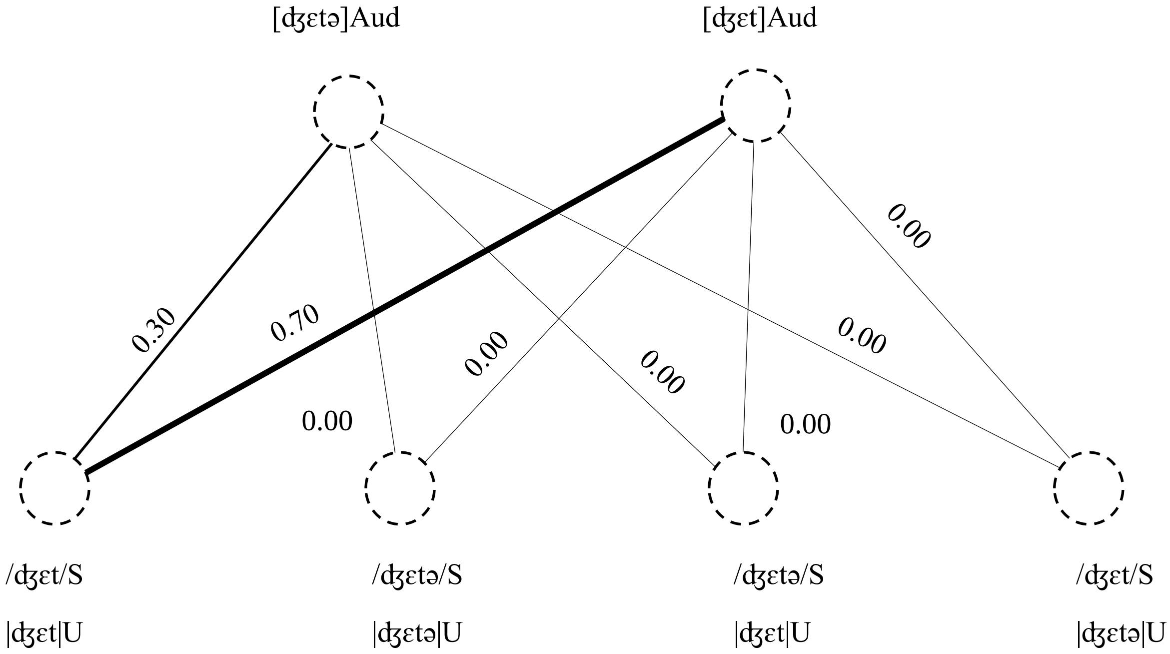

Figure 11. Speaker B's perception of the word 'jet'.

It is in production, however, that the two speakers behave differently. As can be seen in Figure 12, speaker A, which received a low amount English input, retrieves |ʤɛt|U from underlying form, phonologically produces /ʤɛt/S, and phonetically produces [ʤɛt]Aud, a somewhat weaker consonant release, 10% of the time, and produces [ʤɛtə]Aud, a stronger consonant release with the insertion of an IV, 90% of the time. In fact, speakers tend to reflect in production the same degree of variation that they perceived (Boersma 1997). By contrast, speaker B, who has received a high amount of native English input, retrieves |ʤɛt|U from underlying form, /ʤɛt/S in surface form, and phonetically produces [ʤɛt]Aud 70% of the time, and [ʤɛtə]Aud 30% of the time (Figure 13).

[11] Admittedly, with high amount of native English input, a speaker could start to perceive the IV in surface form, provided their acoustic cues for a consonant release have completely shifted toward the English ones. I am assuming here this is not the case, and I will leave the issue to future investigations.

Therefore, I hypothesize that the great quantity of English input received by speaker B has likely changed their acoustic cues for a strong release, making it shift towards the English one.

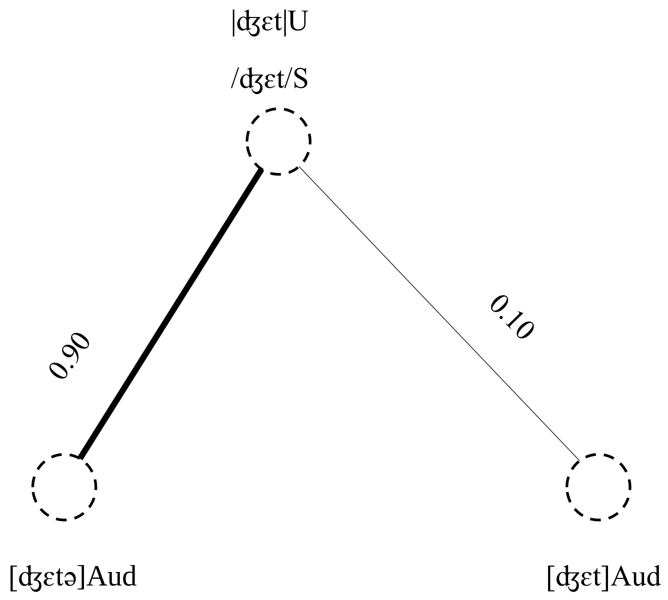

Figure 12. Speaker A's production of the word 'jet'.

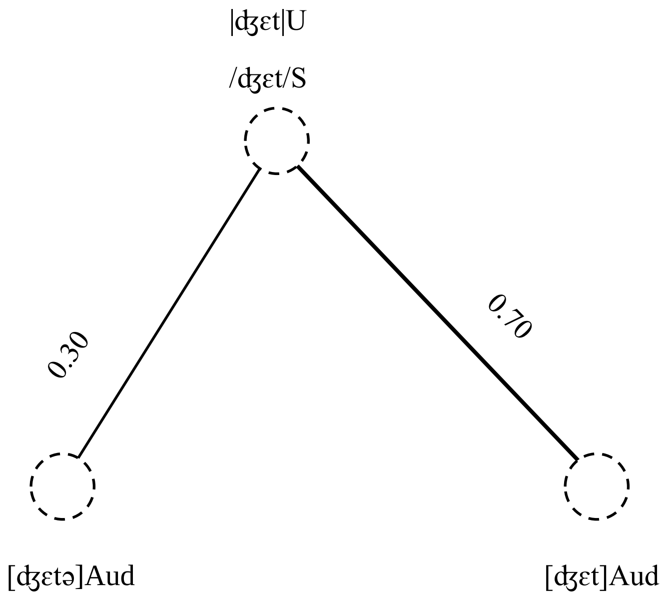

Figure 13. Speaker B's production of the word 'jet'.

5.2 Linguistic factors

In this first analysis, the contrasts were set this way: VDOB = -1/2, +1/3 ; VLOB = +1/2, +1/3; SN = 0, -2/3. A significant result was found for **voicing**, in the comparison between VDOB and VLOB, with the odds of IVs being produced being 10.97 times higher when the consonant is voiced, as opposed to when it is voiceless (95% CI 4.59..26.25 times, $Pr(>|z|) = 3.94*10^{-8}$). Please note that in this analysis not all consonants where compared, but only voiced and voiceless obstruents. On the other hand, the **consonant type** obstruent vs sonorant comparison was done on the whole dataset. Significant results were also found for this predictor, with the odds of IVs being produced being 2.28 times higher when the consonant is sonorant (95% CI 0.998..5.19 times, $Pr(>|z|) = 0.046$). Our prediction was that vowels would be inserted less frequently after sonorants due to influence of Venetian. Opposite to this, we found that vowels are inserted to a greater extent in this context. One explanation for this counterintuitive outcome is that sonorants favor the insertion of IVs because they are voiced. In the model, sonorants are contrasted to obstruents, which include both voiced and voiceless consonants, which would justify the low significance of the $Pr(>|z|)$-value. Another statistical analysis was run to verify this explanation, in which the same model as in (5) was used, although the contrasts for consonant type were changed. In this second analysis, first sonorants and voiced obstruents are compared with each other, and then compared to voiceless obstruents (contrasts: VDOB = -1/2, +1/3 ; VLOB = 0, -2/3; SN = +1/2, +1/3). The results show a non-significant difference between VDOB and SN (estimation = 1.45 times, 95% CI 0.58..3.67 times, $Pr(>|z|) = 0.42$), and a very significant difference between VDOB and SN vs VLOB, with an effect size of 9.09 (95% CI 4.17..20 times, $Pr(>|z|) = 1.54*10^{-8}$), meaning that voiceless obstruents are 9.09 times more likely to be produced with a following IV than sonorants and voiced obstruents. The similar effect sizes in the difference between VLOB and VDOB, and between VDOB and SN vs VLOB, and the fact that the difference between VDOB and SN does not show a significant difference in the same direction as the contrast between SN and OB, suggest that the significant difference between sonorants and obstruents is due to voicing.

Therefore, the results here found confirm those of Grice et al.'s (2015) study, in that voicing is a strong factor in the production of IV insertion. Moreover, their findings stating that type of consonant does not appear to have an independent effect on vowel insertion are not contradicted by our findings.

6. CONCLUSIONS

The findings of this study are twofold. First, we corroborated previous studies such as Grice et al. (2015), which found a strong predictor for the insertion of IVs in the voicing of the preceding consonant. Second, a significant correlation was found between input and the production of IVs, with the latter decreasing as the amount of native English input increases. This last result provides empirical support to stochastic-based models of language learning such as Stochastic OT, NNs, and the Maximum Entropy Model. It was also argued that using input to investigate L2 influence on loanword adaptation is a more direct way than other measurements, and that it is fundamental to include in the estimation input received from passive sources, such as the vision of media content. Finally, we explained how input influences the production of IVs using a BiPhon-NN model.

Using the amount of input as a method to assess the L2 influence over L1 phonotactics has some shortcomings. First and foremost, the real amount of input received is impossible to determine, and it is necessary to resort to estimations. In addition, the way input was here calculated is participant-subjective, and it relied entirely on what they reported. For this reason, it may present some bias, which is unknown to the author.

How input is estimated could also be developed in future studies. For example, we did not include in the study the number of years our participants had been engaged with the vision of

English media in original language. Admittedly, when asked, they had difficulty in remembering, and were the first to acknowledge the fact that their frequency of media vision varied from one period of time to another. This issue will have to be taken into account in future research. Finally, future confirmatory studies will need to explore more in-depth the reliability of L2 input estimation as a predictor of gradience in the adaptation of loanwords.

REFERENCES

Albright, A., & Hayes, B. (2011). Learning and Learnability in Phonology. In Goldsmith, J.A., Riggle, J., & Yu, A.C.L. (Eds.). *The Handbook of Phonological Theory*, 661-690. Oxford, UK: Blackwell.

Almeida, P.A., & Costa, P.D. (2014). Foreign language acquisition: the role of subtitling. *Procedia WCLTA 2013 - Social and Behavioral Sciences, 141*, 1234 – 1238.

Ashton, K. (2014). Using self-assessment to compare learners' reading proficiency in a multilingual assessment framework. *System, 42*, 105-119.

Bafile, L. (2005). Struttura sillabica e consonanti finali in varietà italiana. *Quaderni del Dipartimento di Linguistica - Università di Firenze, 15*, 1-25.

Baker, W. (2005). The Interaction of the Bilingual's Two Phonetic Systems: Differences in Early and Late Korean-English Bilinguals. In Cohen, J., McAlister, K.T., Rolstad, K., & MacSwan, J. (Eds.). *Proceedings of the 4th International Symposium on Bilingualism*, 163-174. Somerville, USA: Cascadilla Press.

Baltazani, M., & Nicolaidis, K. (2013). The many faces of /r/. In Spreafico, L., & Vietti, A. (Eds.). *Rhotics. New data and perspectives*, 125-144. Bolzano, Italy: Bolzano University Press.

Berruto, G. (1989). Main topics and findings in Italian sociolinguistics. *International Journal of the Sociology of Language, 76*, 7-30.

Boersma, P. (1997). How we learn variation, optionality and probability. *Institute of Phonetic Sciences, University of Amsterdam, Proceedings, 21*, 43-58.

Boersma, P. (2007). Some listener-oriented accounts of h-aspiré in French. *Lingua, 117*(12), 1989-2054.

Boersma, P., Benders, T., & Seinhorst, K. (2013). *Neural networks models for phonology and phonetics* (unpublished manuscript). Retrieved on March 3rd, 2018 from http://www.fon.hum.uva.nl/paul/papers/BoeBenSei37.pdf.

Boersma, P., & Hamann, S. (2009a). Loanword adaptation as first-language phonological perception. In Calabrese, A., & Wetzels, W.L. (Eds.). *Loan phonology*, 11-58. Amsterdam, The Netherlands: John Benjamins.

Boersma, P., & Hamann, S. (2009b). Introduction: models of Phonology in perception. In Boersma, P., & Hamann, S. (Eds.). *Phonology in perception*, 1-24. Berlin, Germany: De Gruyter.

Boersma, P., & Hayes, B. (2001). Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry, 32*(1), 45-86.

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer (Version 6.0.37) [Computer program]. Retrieved December 10th, 2017, from http://www.fon.hum.uva.nl/praat/.

Borrás, I., & Lafayette, R.C. (1994). Effects of Multimedia Courseware Subtitling on the Speaking Performance of College Students of French. *The Modern Language Journal, 78*(1), 61-75.

Broniś, O. (2016). Italian vowel paragoge in loanword adaptation: Phonological analysis of the Roman variety of Standard Italian. *Italian Journal of Linguistics, 28*(2), 25-68.

Brown, N.A., Dewey, D.P., & Cox, T.L. (2014). Assessing the Validity of Can‐Do Statements in Retrospective (Then‐Now) Self‐Assessment. *Foreign Language Annals, 47*(2), 261-285.

Cardinaletti, A., & Repetti, L. (2007). Vocali epentetiche nella morfologia dell'italiano e dei dialetti italiani. In Maschi, R., Penello, N., & Rizzolatti, P. (Eds.). *Miscellanea di studi linguistici offerti a Laura Vanelli da amici e allievi padovani*, 116-126. Udine, Italy: Forum Editrice.

Catford, J.C. (2001). On Rs, rhotacism and paleophony. *Journal of the International Phonetic Association*, *31*(2), 171-185.

Coetzee, A., & Pater, J. (2011). The Place of Variation in Phonological Theory. In Goldsmith, J.A., Riggle, J., & Yu, A.C.L. (Eds.). *The Handbook of Phonological Theory*, 661-690. Oxford, UK: Blackwell.

Council of Europe. (2001). *Common European framework of reference for languages: learning, teaching, assessment.* Cambridge, UK: Cambridge University Press.

De Mauro, T. (2018). *Il Nuovo De Mauro*. Available online at https://dizionario.internazionale.it/.

Edele, A., Seuring, J., Kristen, C., & Stanat, P. (2015). Why bother with testing? The validity of immigrants' self-assessed language proficiency. *Social Science Research*, *52*, 99-123.

Flege, J.E. (2007). Language contact in bilingualism: phonetic system interactions. In Cole, J., & Hualdo, J.I. (Eds.). *Laboratory Phonology 9*, 353-381. Berlin, Germany: De Gruyter.

Flege, J.E., Schirru, C., & MacKay, I.R.A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467-491.

Grice, M., Savino, M., Caffò, A., & Roettger, T.B. (2015). The tune drives the text - Schwa in consonant-final loan words in Italian. In The Scottish Consortium for ICPhS 2015 (Ed.). *Proceedings of the 18th International Congress of Phonetic Sciences*, 1-5. Glasgow, UK: the University of Glasgow. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICP HS0464.pdf.

Hall, N. (2006). Cross-linguistic patterns of vowel intrusion. *Phonology*, *23*(3), 387-429.

Hall, N. (2011). Vowel epenthesis. In van Oostendorp, M., Ewen C.J., Hume, E., & Rice, K. (Eds.). *The Blackwell companion to phonology*, 1576-1596. Oxford, UK: Wiley-Blackwell.

Hancin-Bhatt, B., & Bhatt, R.M. (1997). Optimal L2 syllables: Interaction of transfer and developmental effects. *Studies in Second Language Acquisition*, *19*(3), 331-378.

Hayes, B., & Wilson, C. (2008). A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry*, *39*(3), 379-440.

Hayward, K. (2013). *Experimental phonetics*. New York, USA: Routledge.

Kager, R. (1999). *Optimality Theory*. Cambridge, UK: Cambridge University Press.

Kang, Y., & Schertz, J. (2017). *The influence of perceived L2 sound categories in on-line adaptation and implications for loanword phonology* (unpublished manuscript). Retrieved on January 7th, 2018 from https://www.yoonjungkang.com/uploads/1/1/6/2/11625099/kang_schertz_sep2017_not _anonymous.pdf.

Köpke, B., Schmid, M.S., Keijzer, M.C.J., & Dostert, S. (Eds.). (2007). Language Attrition: Theoretical perspectives. Amsterdam, The Netherlands: John Benjamins.

Kwon, H. (2017). Language experience, speech perception and loanword adaptation: Variable adaptation of English word-final plosives into Korean. *Journal of Phonetics*, *60*(1), 1-19.

LeBlanc, R., & Painchaud, G. (1985). Self-Assessment as a Second Language Placement Instrument. *TESOL Quarterly*, *19*(4), 673-687.

Lindau, Mona. (1980). The story of /r/. *The Journal of the Acoustical Society of America*, *67*(S1), S27.

McCarthy, J.J. (2008). *Doing Optimality Theory*. Oxford, UK: Blackwell.

Mistar, J. (2011). A study of the validity and reliability of self-assessment. *TEFLIN Journal, 22*(1), 45-58.

Nespor, M. (1993). *Fonologia*. Bologna, Italy: II Mulino.

Nomura, J., & Ishikawa, K. (2018). Effects of first language processes and representations on second language perception: The case of vowel epenthesis by Japanese speakers. *International Journal of Bilingualism, 22*(1), 69-87.

Passino, D. (2008). *Aspects of consonantal lengthening in Italian*. Padua, Italy: UniPress.

Prince, A., & Smolensky, P. (1993 [2004]). *Optimality Theory: Constraint interaction in Generative Grammar*. Oxford, UK: Blackwell.

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna: Austria. Retrieved on February 2017 from https://www.R-project.org/.

Rabanus, S. (2003). Intonation and syllable structure: a cross-linguistic study of German and Italian conversations. *Zeitschrift für Sprachwissenschaft, 22*(1), 86-122.

Repetti, L. (1993). The integration of foreign loans in the Phonology of Italian. *Italica, 70*(2), 182-196.

Repetti, L. (2006). The emergence of marked structures in the integration of loans in Italian. In Gess, R.S., & Arteaga, D. (Eds.). *Historical Romance Linguistics: Retrospective and Perspectives*, 209-238. Amsterdam, The Netherlands: Benjamins.

Repetti, L. (2009). Gemination in English loans in Americans varieties of Italian. In Calabrese, A., & Wetzels, W.L. (Eds.). *Loan phonology*, 225-240. Amsterdam, The Netherlands: Benjamins.

Repetti, L. (2012). Consonant-final loanwords and epenthetic vowels in Italian. *Catalan Journal of Linguistics, 11*, 167-188.

Rice, M. (1983). The role of television in language acquisition. *Developmental review, 3*(2), 211-224.

Romano, A. (2013). A preliminary contribution to the study of phonetic variation of /r/ in Italian and Italo-Romance. In Spreafico, L., & Vietti, A. (Eds.). *Rhotics. New data and perspectives*, 125-144. Bolzano, Italy: Bolzano University Press.

Ross, S. (1998). Self-assessment in second language testing: a meta-analysis and analysis of experiential factors. *Language Testing, 15*(1), 1-20.

Stuart-Smith, J., Pryce, G., Timmins, C., & Gunter, B. (2013). Television can also be a factor in language change: Evidence from an urban dialect. *Language, 89*(3), 501-536.

Tesar, B. (1997). Multi-recursive Constraint Demotion. *Rutgers Optimality Archive, 126*. Retrieved on June 2018 at http://roa.rutgers.edu/files/197-0597/197-0597-TESAR-0-0.PDF.

Tesar, B, & Smolensky, P. (1998). Learnability in Optimality Theory. *Linguistic Inquiry, 29*(2), 229-268.

Ursini, F. (2011). *Veneti, dialetti*. Retrieved on 4th June from http://www.treccani.it/enciclopedia/dialetti-veneti_(Enciclopedia-dell'Italiano)/.

Wainer, H., Gessaroli, M., & Verdi, M. (2006). Finding What Is Not There through the Unfortunate Binning of Results: The Mendel Effect. *Chance, 19*(1), 49-52.

APPENDIX A

List of words in the irregular plurals elicitation task (words with asterisk are fillers)

Casa (trial)	Link	Autobus	Freezer
Letto (trial)	Braccio*	Stato*	Menù*
Cane (trial)	Jeans	Jeep	Campus
Computer	Bunker	Bulldog	Scarpa*
Uovo*	Città*	Stivale*	Boomerang
Suv	Cactus	Geyser	Mouse
Uomo*	Hotel	Bue*	Dio*
Smoking	Greco*	Macchina*	Flashback
Tunnel	Goal	Kilt	Duca*
Camicia*	Eco*	Cocktail	Residence
Jet	Hostess	Paio*	Aereo*
Sedia*	Risata*	Camper	Scout
Bar	Tram	Caffè*	Scoop
Weekend	Telefono*	Blog	Tazza*
Centinaio*	Tablet	Mano*	Hashtag
Kit	Festival	Club	Ciliegia*
Tempio*	Bagaglio*	Hamburger	Performance
Grido*	Poster	Radio*	Palo*

Handicap	Password	Scooter	Gnu*
Laser	Cioccolatino*	Hacker	Skateboard
Miglio*	Designer	Libro*	Slide
Chat	Musical	Vite*	Piazza*
Wafer	Sciarpa*	Cast	Bob
Bottiglia*	Network	Live	Catering
Ipad	Edificio*	Spazio*	Cesto*
Gossip	Tabloid	Band	Download
Ginocchio*	Piercing	Icerberg	Trend
Desktop	Orecchio*	Sdraio*	Coperta*
Canyon	Leader	Budget	Rollerblade
Ala*	E-mail	Monitor	
Jackpot	scaffale*	Tappeto*	
Strada*	Bowling	Fan	
Pancake	Tappo*	Sito web	

APPENDIX B
Guided questionnaire (English translation)

Part A – general questions

(1) How old are you? _______________

(2) Where are you from (born and raised)? Region: _______________________

(3) Where are your parents from? Region: ___________________________

(4) What are you studying at university? Where? ______________________

(5) What other languages do you know besides Italian? At what level (beginner/intermediate/advanced)?

_______________________ _______________________

_______________________ _______________________

_______________________ _______________________

Part B – contact with foreign languages

(6) In what country (countries) and have you spent more than one month? (name all)

Country: _______________________ for the period of ________________________

Country: _______________________ for the period of _________________________

Country: _______________________ for the period of _________________________

(7) **If you have been in an English-speaking country**, how much did you verbally interact, on average, with English native speakers (friends, teachers, hosting family, local community etc)?

- ☐ More than 6 hours per day
- ☐ Between 2-6 hours per day
- ☐ Less than 2 hours per day

(8) **If you have been in a non-English-speaking country** (other than your home country), how much did you verbally interact, in average, with native speakers (friends, teachers, hosting family, local community etc)?

- More than 6 hours per day
- Between 2-6 hours per day
- Less than 2 hours per day

(9) **If you have been in a non-English-speaking country** (other than your home country), how much did you verbally interact, on average, with **English** native speakers in **English** (friends, teachers, hosting family, etc)?

- More than 6 hours per day
- Between 2-6 hours per day
- Less than 2 hours per day

(10) Do you have any regular contact (verbal interaction, including lectures) with English native speakers in your **home country**?

- Between 5-8 hours per day / week / month
- Between 3-5 hours per day / week / month
- Between 0-3 hours per day /week / month
- Never

(11) Do you regularly watch video / TV shows / movies in English? Yes / No

(12) If you answered question (11) with yes, do you watch them with subtitles? Yes / No

(13) If you answered question (12) with yes, are the subtitles in Italian or in English? It / En

(14) **If you watch videos**, how many hours **per week** do you spend watching videos in English?

- 21 – 30 hours per week (approx. 3-4 hours per day)
- 11 – 20 hours per week (approx. 2-3 hours per day)
- 3 – 10 hours per week (approx. 1-2 hour per day)
- Less than 3 hours per week

(15) **If you watch movies**, how many hours **per week** do you spend watching movies in English?

- 21 – 30 hours per week (approx. 3-4 hours per day)
- 11 – 20 hours per week (approx. 2-3 hours per day)
- 3 – 10 hours per week (approx. 1-2 hour per day)
- Less than 3 hours per week

(16) **If you watch TV shows**, how many hours **per week** do you spend watching TV shows in English?

- ☐ 21 – 30 hours per week (approx. 3-4 hours per day)
- ☐ 11 – 20 hours per week (approx. 2-3 hours per day)
- ☐ 3 – 10 hours per week (approx. 1-2 hour per day)
- ☐ Less than 3 hours per week

(17) Do you regularly come into contact with native speakers of English in other ways? If yes, please specify

____

____

____

____

APPENDIX C
Self-assessment English proficiency test (English version)

Adapted from Council of Europe (2001) .

LISTENING

- I can recognize familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.

- I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.

- I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.

- I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of films in standard dialect.

- I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. I can understand television programmes and films without too much effort.

- I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided I have some time to get familiar with the accent.

READING

- I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.
- I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.

- I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.

- I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.

- I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.

- I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.

SPOKEN INTERACTION

- I can interact in a simple way provided the other person is prepared to repeat or rephrase things at a slower rate of speech and help me formulate what I'm trying to say. I can ask and answer simple questions in areas of immediate need or on very familiar topics.

- I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar topics and activities. I can handle very short social exchanges, even though I can't usually understand enough to keep the conversation going myself.

- I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).

- I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible. I can take an active part in discussion in familiar contexts, accounting for and sustaining my views.

- I can express myself fluently and spontaneously without much obvious searching for expressions. I can use language flexibly and effectively for social and professional purposes. I can formulate ideas and opinions with precision and relate my contribution skillfully to those of other speakers.

- I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms. I can express myself fluently and convey finer shades of meaning precisely. If I do have a problem I can backtrack and restructure around the difficulty so smoothly that other people are hardly aware of it.

SPOKEN PRODUCTION

- I can use simple phrases and sentences to describe where I live and people I know.

- I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my educational background and my present or most recent job.

- I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions.

- I can present clear, detailed descriptions on a wide range of subjects related to my field of interest. I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.

- I can present clear, detailed descriptions of complex subjects integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.

- I can present a clear, smoothly-flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points.

WRITING

- I can write a short, simple postcard, for example sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form.

- I can write short, simple notes and messages relating to matters in areas of immediate needs. I can write a very simple personal letter, for example thanking someone for something.

- I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.

- I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view. I can write letters highlighting the personal significance of events and experiences.

- I can express myself in clear, wellstructured text, expressing points of view at some length. I can write about complex subjects in a letter, an essay or a report, underlining what I consider to be the salient issues. I can select style appropriate to the reader in mind.

- I can write clear, smoothly-flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works.

APPENDIX D
Praat scripts

Table creation

```
speaker$ [1] = "AA_06_M"
speaker$ [2] = "AB_02_F"
speaker$ [3] = "AM_07_F"
speaker$ [4] = "BB_01_F"
speaker$ [5] = "BI_11_F"
speaker$ [6] = "CD_05_F"
speaker$ [7] = "DB_18_F"
speaker$ [8] = "FF_14_F"
speaker$ [9] = "GB_04_M"
speaker$ [10] = "GF_19_F"
speaker$ [11] = "GZ_08_F"
speaker$ [12] = "LC_03_M"
speaker$ [13] = "MD_16_F"
speaker$ [14] = "MR_20_F"
speaker$ [15] = "MZ_09_F"
speaker$ [16] = "PI_13_F"
speaker$ [17] = "PL_15_M"
speaker$ [18] = "SC_21_F"
speaker$ [19] = "SF_12_F"
speaker$ [20] = "SR_17_F"
speaker$ [21] = "SS_10_M"

numberOfSpeakers = 21

writeInfoLine: "speaker", tab$, "gender", tab$, "tone", tab$, "word", tab$, "durationlab", tab$,
"duration"

for i to numberOfSpeakers
	code$ = speaker$ [i]
	genderLetter$ = right$ (code$, 1)
	gender$ = if genderLetter$ = "F" then "female" else "male" fi
	speaker$ = left$ (code$, 5)
	@doFile: code$
	procedure doFile: fileName$
		code2$ = fileName$
		textgrid = Read from file: fileName$ + ".TextGrid"
		numberOfIntervals = Get number of intervals: 2
		numberOfIntervals2 = Get number of intervals: 3
		numberOfIntervals3 = Get number of intervals: 4
		assert numberOfIntervals = numberOfIntervals3
		for interval to numberOfIntervals
			label2$ = Get label of interval: 3, interval
			label$ = Get label of interval: 2, interval
			label3$ = Get label of interval: 4, interval
			if label3$ <> ""
```

```
                start = Get starting point: 4, interval
                end = Get end point: 4, interval
                duration = end - start
                appendInfoLine: code$, tab$, genderLetter$, tab$, label2$, tab$,
label$, tab$, label3$, tab$, fixed$ (duration, 3)
            endif
        endfor
        removeObject: textgrid
    endproc
endfor
```

Presence of vocalic elements

```
speaker$ [1] = "AA_06_M"
speaker$ [2] = "AB_02_F"
speaker$ [3] = "AM_07_F"
speaker$ [4] = "BB_01_F"
speaker$ [5] = "BI_11_F"
speaker$ [6] = "CD_05_F"
speaker$ [7] = "DB_18_F"
speaker$ [8] = "FF_14_F"
speaker$ [9] = "GB_04_M"
speaker$ [10] = "GF_19_F"
speaker$ [11] = "GZ_08_F"
speaker$ [12] = "LC_03_M"
speaker$ [13] = "MD_16_F"
speaker$ [14] = "MR_20_F"
speaker$ [15] = "MZ_09_F"
speaker$ [16] = "PI_13_F"
speaker$ [17] = "PL_15_M"
speaker$ [18] = "SC_21_F"
speaker$ [19] = "SF_12_F"
speaker$ [20] = "SR_17_F"
speaker$ [21] = "SS_10_M"

numberOfSpeakers = 21

writeInfoLine: "EP"

for i to numberOfSpeakers
    code$ = speaker$ [i]
    @doFile: code$
    procedure doFile: fileName$
        code2$ = fileName$
        textgrid = Read from file: fileName$ + ".TextGrid"
        numberOfPoints = Get number of points: 1
        for point to numberOfPoints
            label3$ = Get label of point: 1, point
            if label3$ <> ""
```

```
                appendInfoLine: label3$
            endif
        endfor
        removeObject: textgrid
    endproc
endfor
```

Appendix E

R script

````
```{r}
table <- read.delim ("data/data.txt")
table
```

```{r}
#contrasts

levels (table$proficiency)
contrast <- cbind (c(+1/2, -1/2)) # H = +1/2, L = -1/2
colnames (contrast) <- c("-L+H")
contrasts (table$proficiency) <- contrast

levels (table$type3)
contrast <- cbind (c(-1/2, +1/2, 0), c(+1/3, +1/3, -2/3)) #VDOB = -1/2, +1/3 ; VLOB = +1/2, +1/3; SN = 0, -2/3
colnames (contrast) <- c("-VLOB+VDOB", "-VLOBVDOB+SN")
contrasts (table$type3) <- contrast

table$input <- as.numeric(table$input)
table$inputC <- scale(table$input, center = TRUE, scale = FALSE)
```

```{r}
#model with everything
library (lme4)

model2 <- glmer (EP ~ inputC * proficiency * type3  + (type3 | speaker) + (1 | word), family=binomial, data=table, control = lmerControl (optCtrl = list (maxfun = 100000)))

summary (model2)

#inputC
coefficients = coefficients (summary (model2))
estimate <- coefficients ["inputC", "Estimate"]
se <- coefficients ["inputC", "Std. Error"]
lb <- estimate - 2.0 * se
up <- estimate + 2.0 * se

exp(estimate)
exp(lb)
exp(up)

#proficiency
coefficients = coefficients (summary (model2))
````
````

```
estimate <- coefficients ["proficiency-L+H", "Estimate"]
se <- coefficients ["proficiency-L+H", "Std. Error"]
lb <- estimate - 2.0 * se
up <- estimate + 2.0 * se

exp(estimate)
exp(lb)
exp(up)

#type3- voicing
coefficients = coefficients (summary (model2))
estimate <- coefficients ["type3-VLOB+VDOB", "Estimate"]
se <- coefficients ["type3-VLOB+VDOB", "Std. Error"]
lb <- estimate - 2.0 * se
up <- estimate + 2.0 * se

exp(estimate)
exp(lb)
exp(up)

#type3- type
coefficients = coefficients (summary (model2))
estimate <- coefficients ["type3-VLOBVDOB+SN", "Estimate"]
se <- coefficients ["type3-VLOBVDOB+SN", "Std. Error"]
lb <- estimate - 2.0 * se
up <- estimate + 2.0 * se

exp(estimate)
exp(lb)
exp(up)
```