DOI: 10.17469/O2106AISV000019

DUCCIO PICCARDI

Bilingual Frequency in a Favorable Context (BFFC) in the Italian dialectal area. Theoretical preliminaries to the analysis of geminate lateral retroflexion and voiceless plosives aspiration in Antona (MS)

The paper builds a theoretical framework for the application of the Bilingual Frequency in Favorable Context (BFFC) formula to the peculiar Italian linguistic setting. BFFC was first devised as a usage-based tool to weight the frequency effect of non-varying cognate words against the probability of variation phenomena in bilingual settings. Since Italian dialects are sister languages of the standard variety, speakers can be considered bilingual. However, no dialectal frequency corpora for the extraction of essential BFFC components are available. The paper suggests overcoming this hurdle using subjective frequency estimates and testing BFFC effectiveness through a picture-naming task and acceptability ratings. A critical overview of the phonetic features of interest is also presented, advancing proposals for future analyses.

Key words: Bilingual Frequency in a Favorable Context, retroflexion, aspiration, usage-based linguistics, dialectology.

1. *Two focal points of usage-based theory*

Langacker (1987: 370) was the first to advocate «the necessity for a nonreductive, usage-based model» of linguistic analysis. In recent years, a converging scientific trend (Bybee, 2006) brought together specialists from different areas, such as grammaticalization, variationist sociolinguistics, phonetics and first language acquisition, to examine corpora of natural language use. Here, communicative experience is deemed responsible for speakers' adjustments of gradient boundaries of linguistic categorizations, conditioning production and perception. Usage-based approaches consider language as a part of domain-nonspecific cognitive processes affecting human behaviors. In particular, great attention is dedicated to cooccurrence detection as a function of frequency of exposure, a «domain-general process of the human mind» (Bybee, Beckner, 2009: 830).

Phonetics adopted a usage-based mindset through exemplar models of representation, thanks to the seminal works of theoretical systematization by Johnson (1997), Pierrehumbert (2001, 2002) and Bybee (2001). In this framework, exemplars are stored as eventive traces in the mind of the individual, including fine-grained phonetic details and extralinguistic information (context, participants etc.). These episodic units are structured in metaphorical clouds, organized by the similarity of their components.

In the first section of this paper, two aspects of usage-based approaches will be discussed, namely which phonological phenomena they can account for and which frequency they are about. Albeit central, these topics may still need some degree of clarification. After introducing a challenging setting of research, i.e. the Italian dialectal area, where frequency corpora are not available, I will discuss the potential for this line of research of recent expedients of usage-based bilingual phonetics. Finally, suggestions and solutions will be sketched in a theoretical preview of a research in progress on the production of geminate lateral retroflexion and voiceless plosives aspiration in the dialect of Antona (MS).

1.1 Variation and reduction in frequency effects

The two conditioning forces that token frequency induces on the role of single words in potential sound change are labeled in Bybee's framework Conserving Effect and Reducing Effect (e.g. Bybee, 2006), accounting for analogical and phonetically based shifts, respectively. As I will propose below, this bipartition may not represent the whole spectrum of inherent phenomena; in this vein, Barras, Honeybone & Trousdale (2007) advanced an opposition between Frequency and Anti-Frequency Effects, putting stress on the directionality of frequency tendencies on language change.

As experience of similar linguistic events grows, speakers' categorization of the so formed exemplar cluster strengthens. Central, denser exemplar groups may attract nearby clouds, promoting analogical change in their direction. In the target clouds, marginal (and less frequent) events are more prone to drift towards nearby stronger categories. High-frequency words tend to *conserve* their categorial affiliation, while low-frequency words lead in analogical change. De Schryver, Neijt, Ghesquière & Ernestus (2008) present a rather straightforward instance of this pattern. In a judgement of standardness test, Dutch participants selected as standard the words containing voiceless allophones of phonological voiced fricatives if the stimuli were in the low frequency range, as assessed through subjective estimations (see below, 1.2). This result, modulated by diatopic factors, substantiates the hypothesis of a sound merger in progress between the voiced and voiceless fricative categories in this language.

In contrast to the trans-categorial, relational assumptions required by usage-based interpretations of analogical change, denser clouds may trigger *per se* neuromotor automation, independently of the structure and the qualities of nearby clusters. Bybee (2002: 268-269) explains this phenomenon as a general process of gestural optimization through constant practice, resulting in anticipatory overlaps (assimilation), omissions of unnecessary components (deletion) and reduction in the magnitude of the action (lenition). These phenomena affect first high frequency words, whose exemplar representations rapidly become dominated by reduced variants. Hay, Foulkes (2016) recently wrote a landmark study observing for the first time a reducing effect leading to actual change in a real time corpus. In New Zealand English, intervocalic medial /t/ sonorization spread following a positive

frequency pattern and interacting with the year of birth of the recorded speaker. This result is fundamental for usage-based phonetics, since it anchors to real diachronic observations what was until then limited to predictions of change or speculations on synchronic data.

In the last years, a sociophonetic line of research, pioneered by Hay, Jannedy & Mendoza-Denton (1999), followed Pierrehumbert (2002: 118)'s assertion that word frequency positively affects the rate of change in the composition of any exemplar cloud, favoring newer variants in overcoming the representational balance: in other words, any type of change, and not only reduction, is led by frequent words. This can be interpreted as a major departure from Bybee's point of view, since frequency effects are not considered as unambiguous flags of cognitively determined change in production[1]. In Pierrehumbert's framework, the explanation of variation and change phenomena remains outside the basic assumptions of the usage-based framework *per se*, i.e. the existence of general cross-domain cognitive mechanisms[2]. Even though the source of variation is ultimately physiological, an explanatory proposal should account for the systemic strategies of acceptance and accommodation of the new reduced variants (Todd, Pierrehumbert & Hay, 2019).

Despite these assumptions, a subtle sense of uneasiness still stands in phonetic research in leaving aside Bybee's approach, usage-based inquiries on strengthening phenomena being quite rare. Moreover, under certain circumstances, variation has been reconducted to reduction through phonological arguments. For example, Habib (2011) framed the lexically-driven, frequency affected change from Standard Arabic /θ ð/ to Hims dialectal /t s/ in a reduction process; analogous choices are made by Coetzee and Kawahara (2013) in their explanation of positive frequency patterns in variable geminate plosive devoicing of English loanwords in Japanese. These authors equate loss of phonological distinctiveness with reduction; however, Bybee's reducing effect should not be extended to phonological domains, these being pertinent to relational, trans-categorial properties. Otherwise, (positive) frequency effects risk becoming an ambiguous and potentially confusing explanatory *passe-partout*. In this sense, note that a robust tradition of usage-based studies (e.g. Brown, 2018 and references therein) applies to plosive spirantization the same principles evoked in Habib (2011) for a change in the exact opposite direction.

Last in this paragraph, it is noteworthy that usage-based theories are conceived to provide an explanation to language change, but not as much to stable variation. For example, it has been pointed out that other models, like Lindblom's (1990) H&H Theory, can easily explain stable reductions principally targeting high-frequency words (Forrest, 2017: 142) without recurring to exemplarist arguments. While, denying the pertinence of usage-based linguistics to diachrony, Dinkin

[1] An early discussion on frequency effects on non-reducing change is presented in Phillips' works, e.g. (2006: 41-56).

[2] From a sociophonetic viewpoint, Clopper, Mitsch and Tamati (2017: 39) asserted that «just as easy contexts for the listener allow for gestural economy leading to reduction, easy contexts for the listener allow for more extreme social indexing without sacrificing communication».

(2008: 105) defined synchronic reduction as «the real effect of word frequency», Abramowicz (2007) failed in finding frequency patterns in the realizations of reduced (ING) variants. Forrest (2017) clarified the issue, suggesting that researchers should leave aside raw frequencies and look for proofs of sensitivity to word frequencies in contexts promoting variation: if a word is frequently encountered in such environments, we can expect that traces of variation will also be present in the absence of them. Indeed, Forrest found that frequency in specific grammatical and phonological environments did condition (ING) stable reduction[3], strongly promoting a complexification of the frequency variable, which should include the negative effects of non-varying exemplars in order to boost the accuracy of the models.

1.2 Conceptualization of frequency: from objectivity to subjectivity

Usage-based linguistics is constitutionally bound to the idea that individual lives exert a great pressure on the observed data. The formation of linguistic categories and patterns of use are the ultimate result of «a personal odyssey» (Foulkes, 2010: 19), modulating experimental acquisitions. Drager (2011) adopted an objective, direct and individual concept of frequency analyzing the interviews recorded during one year of ethnographic fieldwork in a New Zealand high school. Here, the more a specific function of *like* occurred in the speech of individual students, the more this specific *like* became target of phonetic reductions. Through her extensive ethnographic effort, Drager described an ideal method for usage-based research, which is able to account for individual behaviors without resorting to external sources. Even so, the author herself (ibid.: 697) admits that «much larger corpora would be required to calculate representative speaker token frequencies».

In most of the other cases, practicality drove usage-based research towards the assumption of the so-called *from-corpus-to-cognition-principle*, i.e. the idea that «frequency in text instantiates entrenchment in the cognitive system» (see Schmid, 2000: 38-40). Word representativeness in a corpus is seen as a window on systemic and, ultimately, cognitive structures. As we saw at the end of the previous paragraph, this objective, indirect and communal take on frequency is undergoing a process of refinement, leaving behind raw corpus values. In Bybee (2002), the scholar assumed that, since the exemplar clusters are activated by phonetic strings, cognitive representations do not involve words taken from a phonetically abstract co-text, but instead occur in acoustic sequences which are not equally prone to variation. Her point was substantially left unrefined until the works by Esther Brown[4]. In her Doctoral dissertation, Brown (2004) investigated the variable phenomenon of Spanish initial /s-/ reduction using a spoken corpus from New Mexico and Southern Colorado. She devised a formula, the Frequency in Favorable Context (FFC), which calculates a score dividing the frequency of a word occurring in a context of potential

[3] Analogous results are reported in Raymond, Brown & Healy (2016) for final /t d/ deletion.
[4] For an alternative operationalization, see Eddington, Channer (2010).

variation[5] by its overall occurrences. This score proved itself to be predictive of the effective reduction probability of /s-/ words. Raymond, Brown (2012) further promoted the validity of the method using logistic regression to discern the significance of FFC values from other cooccurring factors, including overall frequencies. Brown, Raymond (2012) attempted to apply FFC to a diachronic analysis. They observed that weighted values, extracted from a large Medieval Spanish text, outscored raw frequencies with regard to the prediction of the [h] outcomes of Latin initial F- (e.g. *hablar* < FABULARI). Note that in this study FFC was the only parameter reaching significance, differently from absolute measures and word transmission history coding (i.e. oral vs. non-oral). Thus, this research decisively defined proofs by which the use of raw textual frequencies should be avoided. Brown (2013) replicated this analysis, along with that on /s/ reduction, while adding a part-of-speech tagging to the statistical model. Verbs held the highest FFC values, and consequently became the most reduced category. The study tries to support the idea that different involvement in variation and change of distinct grammatical categories is just an epiphenomenon of peculiar patterns of use. Lastly[6], Brown (2018) used word FFC ratings to predict the type and the extent of Spanish initial /d/ reduction elicited in a list of words reading task administered to 18 speakers from disparate Spanish-speaking countries. This format was chosen with the aim of reducing the potential conditioning factor of style. FFC ratings significantly correlated not only with the type of elicited variant (low FFC: [t], [d], [nd] vs. high FFC: [ð], [∅]), but also with the intra-categorial degree of articulatory strength, calculated as the intensity difference between the consonant and the following vowel.

Alongside these developments in corpus-driven methods, a subjective, indirect and individual route is tentatively followed by few researchers. To the best of my knowledge, De Schryver *et al*. (2008) was the first usage-based phonetic study to rely on subjective frequency estimates instead of objective, corpus-extracted ones[7]. Word subjective frequency estimation (or word familiarity[8]) is a major field of interest in cognitive research, pursued since shortly after the compilation of the first textual word frequency corpus (Thorndike, Lorge, 1944; see Howes, 1954) through a variety of experimental designs. Among those, a few influential studies between the end of the Seventies and the beginning of the Eighties popularized rating scales (e.g. Toglia, Battig, 1978; Nusbaum, Pisoni, & Davis, 1984). Overall, this line of research

[5] In this specific case, the context favoring variation is the presence of contiguous non-high vowels.

[6] For Forrest (2017) and Raymond, Brown & Healy (2016) see above, 1.1.

[7] After De Schryver *et al*. (2008), subjective measures have been used in phonetic research by Myers, Li (2009) on syllable contraction and tonal merger in Southern Min; Lee, Kapatsinski (2014) and Song, Dalola (2019) on loss of productivity of Korean /n/ insertion; Nadeu (2016) on the reduction of full vowels in unstressed position in Central Catalan compounds; Zhang, Meng (2016) and Yan (2018) on the variable realization of Shanghai Wu tone sandhi and, lastly, by Yan, Zhang (2017) and Yan (2018) on the same phenomenon in Wuxi Wu.

[8] Here I cluster together these two definitions for practicality; however, it should be kept in mind that some differences between the two concepts have been underpinned, especially from the procedural point of view (Balota, Pilotti & Cortese, 2001).

finds significant correspondences between subjects' perceived word frequencies and corpus data, not without some discrepancies (see e.g. Carroll, 1971, Thompson, Desrochers, 2009). For this reason, Ringeling (1984) argued that subjective protocols may be reliable substitutes for textual corpora when the latter are not available at all. Turning back to linguistic inquiries, Balota *et al*. (2001) succeeded in proposing a standardization of the research protocol which has been thereafter adopted by the majority of usage-based phonetic studies dealing with subjective measures. Balota and colleagues collected frequency estimations for 2938 English monosyllabic words from a total of 2254 respondents, using a 7-point Likert scale with labels assigned at each point. This study was probably accepted by the linguistic community because of its unprecedented quantity of observations; moreover, some of its methodological features, such as the insertion of age and diamesic variables, are in line with variationist concerns.

Subjective frequency estimation represents an attractive middle-ground between the time-consuming attention to the individual of Drager (2011) and the rapid, objective corpus observations. In the next chapter I will introduce an ideal ground for the application of this methodology: usage-based dialectology.

2. *Usage-based dialectology*

Anderwald and Szmrecsanyi (2009: 1126) ironically stated that, «in contrast to sociolinguistics, dialectology and corpus linguistics have been rather uneasy bedfellows until relatively recently». This is due to the general scarcity of publicly available corpora of local varieties. Even though crowdsourcing technology is gradually trying to change this picture (e.g. Cenceschi, Trivilini, Sbattella & Tedesco, 2019), for the time being, usage-based dialectology has found very limited expressions. For example, works by Lynn Clark are inserted in a usage-based framework thanks to the author's corpus of field materials (e.g. Clark, 2008); in other occasions, arguments of English dialectology are inferred from a geo-referenced subset of a corpus of the national language (Clark, Watson, 2011). This procedure cannot be exported to the Italian linguistic landscape, where dialects are not varieties of the standard, but standalone languages. In this chapter, the bilingual dynamics of Italian dialect-standard interactions will be discussed; then, methods and resources of usage-based bilingual phonetics will be explored and tentatively adjusted to fit the Italian situation.

2.1 Italian dialects: advergence in bilingualism

In Italy, dialects are «'sisters' of Italian» (Maiden, Parry, 1997: 7), and not local varieties of the standard language. This is no mere semantic subtlety (Regis, 2013), as the Italian acceptation encodes the peculiar history behind its linguistic landscape. In fact, Italian primary dialects[9] (Coseriu, 1980) are direct developments of the Latin

[9] From this point on, I will refer to this category of idioms as "dialect". On the other hand, here "standard" does not stand for the Italian reference linguistic construct, but for its local instantiations (unless

spoken in the regions of the peninsula. After the linguistic debates of the XVI century these varieties came in an asymmetric sociolinguistic contact with a reference prestige variety based on a Florentine koine (e.g. Regis, 2017). After the achievement of political unity and the diffusion of a uniform educational system (Sobrero, 1996), this reference lect was codified as the Italian standard. In the meantime, the onset of intermediate varieties favored the formation of a sociolinguistic continuum, explaining the acceptation of "dialectology" as, *de facto*, the study of systems historically independent from the national one. In addition to that, the structural distance between the national (and regional) standard and the dialects «justif[ies] treating the Italian situation as bilingual rather than merely bidialectal» (Berruto, 1997a: 394)[10]. Note that both recent neurolinguistic (Schmitt, Auer & Ferst, 2019) and psycholinguistic (e.g. Mello, Miozzo & Peressotti, 2017) studies confirmed that this theoretical standpoint reflects an observable cognitive reality, since experimental data from standard-dialect bilectal subjects are comparable with those from bilingual ones.

Recent surveys promoted by the National Institute of Statistics (e.g. ISTAT, 2017), based on people's self-evaluations, highlighted that, while monodialectal speakers are decreasing, bilingualism is rather stable: on this basis, Dal Negro, Vietti (2011: 89) asserted that «on a national level the bilingual mode appears as the unmarked or default choice». This widespread situation of multiple competences led to the loss of functional compartmentalization of the concurring codes, which are often coexisting at the same diaphasic level ("dilalìa" in the terminology of Berruto, 1987). As a consequence, dialects are gradually losing peculiar lexical and phonological features (e.g. Cerruti, 2016). Berruto (e.g. 1989, 1997b) repeatedly manifested dissatisfaction with the use of the term convergence (Sanga, 1985) describing this phenomenon. In fact, the two main prerequisites of convergence (bilateral participation in a centripetal narrowing and sociolinguistically balanced status) are not met in the Italian milieu. For this reason, the scholar (Berruto, 2005a) strongly promoted the adoption of the term advergence (Mattheier, 1996), defining a unilateral movement of the dialect towards an (idealized) standard.

Given dilalìa, it has been argued that one of the most representative loci of advergence is code-switching[11] (Cerruti, Regis, 2005: 198). Studies on this phenomenon noted an extreme variability and compositional freedom in its Italian manifestations. In this sense, Berruto (2011) asserted that, during code-mixing, speakers do not activate one of two competing grammars, nor a permeation of the two, but both grammars at the same time, in a cumulative, unified fashion. Overall, code-switching can be considered a surface phenomenon, which is better described by variationist

otherwise stated). In the words of Berruto (2018: 510), «nobody in Italy can be considered to be a true native speaker of standard Italian. There are, instead, several regional standardized varieties of Italian».

[10] This assumption does not stand for the speakers of Tuscan varieties, whose competence can be better described with the tools of variationist sociolinguistics (or sociophonetics: see the overview in Calamai, 2017). The same can be said for the Roman speakers, since their dialect incurred in a profound tuscanization during the XV, and mostly XVI, centuries (Trifone, 1992: 28-50).

[11] Here I refer to code-switching as a hypernym of all the phenomena of code alternation and mixture, excluding explicit borrowings.

sociolinguistic accounts[12]. Indeed, recent attempts at explaining specific instances of dialect-standard switching (Cerruti, 2018) implemented a probabilistic view on their emergence, akin to comparative variationist methodologies (Tagliamonte, 2013). This theoretical shift somehow reflects the birth of a usage-based approach to code-switching at the international level (Backus, 2015). In this framework, insertional code-switching and borrowings are seen as residing on a single continuum of cognitive entrenchment. Therefore, researchers should compare the frequencies of the alternating structures in order to predict their degree of fixated production as elements of discourse variation promoting eventual innovation. Considering linguistic units larger than the word this approach has not yet been applied to phonetic features. On a general note, studies on the phonetics of code-switching are still very scarce (see e.g. Muldner, Hoiting, Sanger, Blumenfeld & Toivonen, 2019); however, the point here is, even interpreting code-switching as a variation phenomenon occurring in a single, cumulative grammar, we cannot assess the predictive power of individual lexical items on phonetic realizations if we cannot be sure that the switched slot is going to be replaced with a semantically and phonetically similar element. In other words, if the two switching elements do not share the same etymological base, variation will necessarily occur at the lexical, and not phonetic level.

Since Clyne (1967), it has been noted that cognate[13] words may favor the emergence of code-switching. This triggering effect has been repeatedly reported in the Italian context (Alfonzetti, 1992: 240; Berruto, 2004: 65-67; Cerruti, Regis, 2005: 197). More recently, psycholinguistic research on cognates (Broersma, De Bot, 2006; Broersma, Carter, Donnelly & Konopka, 2019) classified them as cognitive nodes, linking separately grouped, but nevertheless coexisting language representations. During bilingual discourse, both grammars are activated at some level; the occurrence of a cognate bolsters the arousal of the recessive code, promoting a switch in its proximity. Interestingly enough, phonetic discrepancies between the two cognate words do not hinder their role as cognitive bridges in the mind of bilinguals (Broersma, 2009). This has also been confirmed for the Italian situation, as real-time studies (Ghimenton, 2013a, 2013b) noted that standard-dialect cognates hold a pivotal role in children's dialect acquisition through minimal inputs.

Wrapping up this paragraph, I would argue that the assessment of the probabilistic emergence of these cognate nodes as phonetically attributable to one of the two coexistent systems may be a viable route to quantify phonetic advergence from a usage-based dialectological perspective.

2.2 The usage-based contribution to cognate analysis

Today, tentative quantifications of phonetic advergence are still rarities, all concerning cognate analysis. Among those, Auer, Schwarz (2014) is explicitly interested

[12] Such approach preludes to a transformation of code-switching in a variation phenomenon of the sociolinguistically dominant language *tout-court* (see Mioni, 1992: 18 in Cerruti, Regis, 2005: 200).
[13] In this work, I will use this term in its psycholinguistic acceptation, implying a degree of formal and semantic similarity between the words sharing the cognate status.

in frequency effects. The authors searched for lexical gradience in Alemannic dialect-standard German vocalic advergence through a corpus subset of spoken interviews held in Baden-Württemberg. Production data are examined in the light of values extracted from a frequency dictionary based on southwest German materials; however, all the conclusions drawn for this analogical drift towards standard vocalic categories are lacking an appropriate quantification of the attracting force, i.e. the clusters of standard German-like productions.

Recently, usage-based phonetics has developed a valuable tool in order to investigate frequency effects in bilingual settings. Brown, Harper (2009) studied the conditioning factors of Spanish final sibilant reduction in two corpora of spontaneous speech, one from New Mexico, the other from Chihuahua. They devised two binary variables, i.e. the English Exemplar Connection (EEC) and the cognate factor, indicating different levels of similarity between the inquired Spanish words and their English equivalents. In the New Mexican data, gathered from bilingual speakers, the stronger are the Spanish word exemplar connections with their English equivalents without the inquired variation phenomenon, the less probable is their reduction. This is explainable by positing the storage of exemplars from the two different languages in single clouds, variably populated by non-reduced English exemplars. On the contrary, EEC and the cognate factor were not predictors of resistance to reduction in the data from Chihuahua (Spanish monolinguals). Brown (2015), building on these results, refined the above-discussed FFC formula, implementing the weighting of cognate exemplars. She examined Spanish /d-/ reduction in a Spanish-English code-switching corpus from New Mexico. First, the methodology of Brown, Harper (2009) was replicated, coding a binary cognate variable (e.g. Sp. *Diferencia*, Eng. *Difference* vs. Sp. *Domingo*, Eng. *Sunday*) while adding FFC[14] to the analysis. Both factors proved to be significant; however, focusing the analysis on the cognate subgroup, FFC lost its predictive power. With the aim of providing a more fine-grained formalization of the effects of the bilingual exemplar clouds, Brown repeated this test substituting FFC with the Bilingual Frequency in Favorable Context (BFFC). BFFC adds the overall (external) corpus frequency of the English cognate words to the divisor, contributing to counter the contextual, variation-favoring effects. Indeed, BFFC succeeded in modeling the differences within the cognate category; moreover, because of its own structure, the new formula accounted for the effects of the binary cognate variable, evidently providing a more compact procedure for this kind of research.

To date, no other attempts have been made at validating BFFC in other bilingual environments[15]. Nonetheless, I would advance that BFFC could be an invaluable

[14] Both values composing the FFC, i.e. the overall word frequency and the number of word occurrence in reducing contexts, were extracted from an external reference oral corpus. Spanish /d-/ reduction is triggered in non-post-pause, post-lateral and post-nasal contexts.

[15] Brown, Amengual (2015) investigated the same phenomenon in a corpus of Spanish-English Puerto Rican bilinguals as well as Spanish plosive Voice Onset Time (VOT) lengthening in the speech of Texan heritage speakers, while adopting the methodology of Brown, Harper (2009).

parameter for the studies on standard-dialect advergence. In fact, note that BFFC structure can account for both frequency effects and anti-frequency effects, interpreted as analogical variations in the direction of dominant standard representations. If BFFC positively correlates with the probability of occurrence of the inquired dialectal variation phenomenon, this should be classified as endogenous, i.e. promoted by dialectal entrenchment and disfavored by competing standard cognates. On the contrary, if BFFC and variation are negatively proportional, we may be observing an analogical shift towards standard norms, i.e. an *advergence*. Moreover, in both cases, comparing results for BFFC and FFC may contribute to shedding some light on the degree of functional overlapping of the two coexisting codes: a dense, dilalic code-switching milieu should lead to robust inter-language exemplar clustering, thus favoring BFFC. In the next section, a protocol for a usage-based phonetic analysis of peculiar dialectal features will be outlined, summing up the potential benefits of subjective estimates and bilingual frequency expedients.

3. *Naming pictures in the Apuan Alps. An Italian BFFC protocol*

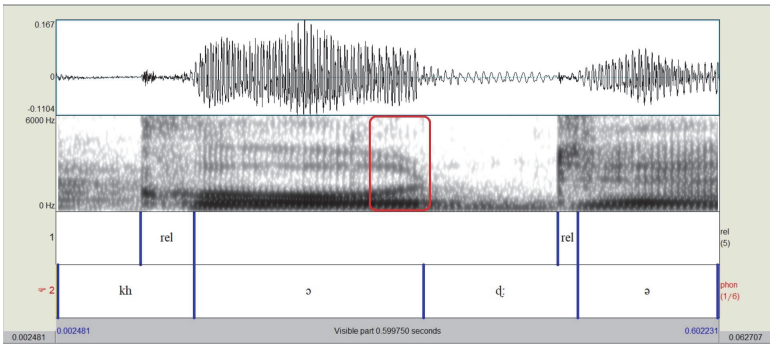
Antona (402 MAMSL; ca. 300 inhab.) is a small village in the upper Frigido valley on the Apuan Alps. Administratively speaking, it is part of the municipality of Massa, a regional chief town in northern Tuscany. The classification of the Antonese dialect is no easy task, as it belongs to the transitional Apuan group (Carpitelli, 1995: 44-47). While Pellegrini (1980^{2}: 30) in his Map of the Italian Dialects placed the Apuan varieties in the Tuscan cluster, Giannarelli (1912: 12) and Ambrosi (1966/1967: 126) considered them a subgroup of the Lunigianese area. Emblematically, both the volumes on Tuscany (Giannelli, 2000: 129-130) and Lunigiana (Maffei Bellucci, 1977: 18-20) which form part of a seminal series on Italian dialects directed by Manlio Cortelazzo hold the other party responsible for discussing the Apuan varieties, ultimately leaving the Italian tradition without a systematic overview on them. In actual fact, as explicated in Giannarelli's (1912) proposal, these dialects are made of an intricate superimposition of Tuscan and Ligurian isoglosses, not without Emilian echoes. Despite the relative isolatedness of the Apuan communities, their advergence to standard features have been noticed long ago (Ambrosi, 1956: 33); conversely, local identity focuses on specific dialect traits, promoting occasional, irregular overextensions (i.e. hyperdialectism: Carpitelli, 2017: 219). On the basis of a general structural contiguity with the Tuscan varieties, I would stress that usage-based accounts could clarify the effects of standard pressure on specific phonetic characteristics with different degrees of local salience. In the remaining space, I will discuss two Antonese features of potential interest, namely geminate lateral retroflexion and voiceless plosives aspiration; then, I will outline a usage-based research in progress on them.

3.1 Retroflexion and aspiration in Antona

Geminate lateral retroflexion (e.g. [ˈvaɖːə] vs. It. [ˈvalːe], "valley") was first reported for the area in Bottiglioni (1911) and its geographic boundaries, revolving around mount Pisanino, were later defined in Bonin (1952: 146) and Ambrosi (1956). This feature was the object of an intense linguistic debate on its origins, variably hypothesized as Corsican (Bottiglioni, 1926: 6), Ligurian (Bottiglioni, 1930: 18) and Mediterranean (Merlo, 1956), until Savoia (1980: 280 and ff.) put an end to the substratist speculations by advancing solid chronological arguments in favor of the relative recency of the trait. Be that as it may, it is undeniable that the singularity of retroflexion is strongly perceived by local speakers: in a description of the Antonese variety written by a non-specialist (Della Sala, 1986), retroflexion is emblematically put first and classified as typical of the most authentic stratum of the population (i.e. elderly people with limited contacts outside the village).

To date, no acoustic data has been published regarding Antonese retroflex plosives. The only spectrographic (and palatographic) information on the phenomenon in this area is reported in Rossi (1976: 415-416) for the variety of Bosco di Rossano in Lunigiana, where, however, the local realization is the approximant [ɭ] and not the plosive [ɖ]. While studies on analogous retroflexions in other Italian dialects found hints towards their classification as strengthening (i.e. long release phases, gradient loss of formant information; for Sicily and Corsica: Celata, 2006), the directionality of variation is irrelevant from a usage-based point of view: in fact, the application of BFFC to the Antonese dialect would encode the gradient nature of an analogical shift towards the Italian concurrent realization, i.e. [lː].

Figure 1 - *[ˈkʰɔɖːə] (It. [ˈkɔlːa], "glue") uttered in continuous speech by a 75 years old female Antonese speaker. Velar release is 49 ms. long, well above the arbitrary 30 ms. threshold set by Cho, Ladefoged (1999: 223) for aspirated plosives. The red box highlights the descending F3/F4 transitions to the retroflex, leading from vowel midpoint 2845, 3848 Hz to endpoint 2087, 2777 Hz, respectively*



On the other hand, a strictly strengthening, dialect-internal variation in Antonese is voiceless plosives aspiration, a phenomenon brought to light in Giannelli, Savoia (1979-1980: 84). Aspiration is reported as mainly targeting velar and dental seg-

ments, both short and long; however, in singleton, intervocalic plosives, aspiration competes with sporadic sonorization, implying, for example, the coexistence of the forms [ˈfoːkʰə] and [ˈfoːg̊ə] for "fire". Again, no acoustic data are available for Antonese aspiration. In Savoia (1980: 242-243) other examples are discussed, and aspiration is synthesized in a variable rule conditioned by social factors, without further specifications. Compared to retroflexion, aspiration is far less salient for local speakers, and is not mentioned in Della Sala (1986). Since the voiceless plosives of standard Italian are realized as unaspirated (e.g. Sorianello, 1996), Antonese has at its disposal both standard-like, unaspirated variants and divergent, aspirated ones. Assuming that positive frequency effects can account for variations other than reduction (cf. Pierrehumbert, 2002), aspiration of Antonese long voiceless plosives should follow cognitive patterns similar to those encoded by the BFFC formula in Brown (2015). Italian exemplars should enhance the density of the unaspirated dialectal ones, countering the emergence of aspirated allophones. Figure 1 illustrates both Antonese aspiration and retroflexion, while relevant acoustic parameters will be mentioned in the next paragraph.

3.2 Subjective/Objective Bilingual Frequency in Favorable Context (SOBFFC) and its applications

This section contains a summary of the current status and future perspectives of a research in progress in Antona, aimed at testing the effectiveness of a usage-based approach centered on BFFC in the Italian bilingual situation.

A list of Italian words containing geminate laterals and voiceless plosives was extracted from itWaC (Baroni, Bernardini, Ferraresi & Zanchetta, 2009) through a case-unsensitive query. In order to slim it down, the words were administered to local informants (two men aged 25 and 30, and a woman aged 56) asking them to gray out those items whose dialect equivalents were hugely dissimilar (i.e., among others, *parpagòn* vs. It. *pipistrello* for "bat"; cf. Mosti, Nancesi, 2005: 378). The remaining words were provisionally[16] considered cognates of dialect forms, whose exact pronunciation was asked as well during this phase. The corpus frequencies of these Italian words were extracted for the BFFC computation and log-transformed (Baayen, 2001: 32-34), while values for their dialect counterparts will be retrieved as follows.

Antonese has a small written tradition, constituted by collections of prayers (Mariotti Giromella, 2006), proverbs and sayings (e.g. Bertuccelli, Casotti, 1982: 43-44). These works are widely known among the villagers: even though extracting

[16] As an anonymous reviewer rightly pointed out, speakers' age plays a major role in determining individual knowledge of the dialect lexicon (e.g. Wieling, Montemagni, Nerbonne & Baayen, 2014). For this reason, the absence of non-cognate alternatives for the extracted Italian words cannot be ascertained without a socially stratified lexicographic inquiry in the village. I should clarify that this was not the goal of this phase. While I was indeed trying to numerically reduce the list of items, the eventual presence of residual non-cognates can be spotted through the results of the rating scales and participants' comments (see below).

from them word frequency values is clearly unfeasible because of their smallness and lack of thematic variety, they created some level of orthographic agreement for non-standard phones, exploitable in testing conditions. The dialect component of BFFC will be gathered through subjective frequency estimation, adapting the guidelines of Balota *et al.* (2001; see above, 1.2). In particular, I will neglect their diamesic bipartition, since written dialect occurrences are still limited to the rare above-mentioned circumstances; moreover, a complexification of their focus on social differences will be needed, adding a sex variable and a classification based on attitudinal criteria. This will be calculated mixing scores taken from three short questionnaires, i.e. the Regionality Index, the Language Use Index and the Ethnic Orientation Index (Dollinger, 2015: 285-294). While the eventual fluctuation in the perception of frequencies depending on dialect attitudinal scores will be object of interest *per se*, these measures will also be exploited for the interpretation of the results of the testing phase. Dialect words will be presented in their orthographic adaptations. This methodological choice will inevitably lead to some level of individual reinterpretation; however, since Myers, Li (2009) proved subjects' sensibility to phonetic details in subjective frequency assessments, proposing acoustic stimuli taken from *ad hoc* elicitations of local informants would have resulted in a large number of uncontrollable factors derived from potential idiolects.

FFC formulae are concerned with weighting the frequency of exposure to context promoting variation against word overall frequency. Note that both retroflexion and aspiration could be analyzed from this point of view. In fact, retroflexion may occur also in across-boundaries contexts, triggered by syntactic doubling (Loporcaro, 1997) in specific word combinations (e.g. [aˈɖːɛtːə] vs. It. [aˈlːɛtːo] "to bed"; Savoia, 1980: 284-285); on the other hand, we saw that singleton voiceless plosives in intervocalic context can either become sonorized or aspirated, depending on social factors. In order to avoid difficulties in representing compositionality and vague extralinguistic coordinates in subjective estimations, the target dialect words will present the inquired phenomena in unambiguous contexts, namely word-internal for retroflexion and in geminates for the aspiration of voiceless plosives, as already coded in the itWaC queries for their standard cognates. By doing so, BFFC will be reformulated as the population mean of the word subjective frequency estimate divided by itself plus the objective frequency of its Italian cognate. I refer to this revised measure as Subjective/Objective Bilingual Frequency in Favorable Context (SOBFFC). Figure 2 illustrates the transition from FFC to SOBFFC.

Figure 2 - *The FFC family of formulae. These equations are synthetic representations of methodological choices adopted in Brown (2004) for FFC, in Brown (2015) for BFFC and in this study for SOBFFC*

$$FFC = \frac{frequency\_in\_variating\_contexts}{overall\_frequency}$$

$$BFFC = \frac{Spanish\_frequency\_in\_variating\_contexts}{Spanish\_overall\_frequency + English\_cognate\_overall\_frequency}$$

$$SOBFFC = \frac{dialect\_mean\_subjective\_frequency}{dialect\_mean\_subjective\_frequency + Italian\_cognate\_overall\_frequency}$$

SOBFFC values should positively correlate with the probability of one word to be produced in its dialect form. Two psycholinguistic methodologies will be exploited in a second session of data collection in order to obtain indirect measures of use. Picture naming tasks are a common procedure for testing the effects of cognitive representations of cognates. Reaction times are usually interpreted as an approximation of the ease in word retrieval which can either be enhanced or impeded by the existence of cognates. However, both in single language (e.g. Hoshino, Kroll, 2008) and language-switching designs (Broersma, Carter & Acheson, 2016) cognate pairs are matched not only by their frequencies, but also by other semantic (e.g. imageability) and lexical (e.g. age of acquisition) properties, which condition the speed of retrieval (e.g. Alario, Ferrand, Laganaro, New, Frauenfelder, & Segui, 2004). While normed corpora of images for Italian picture naming tasks do exist (e.g. Navarrete, Arcara, Mondini & Penolazzi, 2019), the same cannot be said for dialects. Since our inquiry is based on potential attrition between lexical parameters (such as frequency), assuming that the Italian values are valid also for the dialects could be counterproductive. For the time being, at the cost of granularity of observation, we should refrain from testing the processes involved in lexical retrieval and limit our analysis to the contents of the retrieval itself. In other words, subjects will be asked to name the pictures representing the selected words in a simulated context generating potential task ambiguity, such as "in Antona". Through this specification, I intend to stimulate exemplars coded for the communal practices of the village, for which SOBFFC should simulate the degree of dialect advergence to the standard variety. By doing so, the effect of SOBFFC on the rate of retroflex and aspirated responses will be evaluated.

Aspiration can be represented as a temporal interval between the release of a plosive and the beginning of signal periodicity of the following vowel (i.e. VOT: Lisker, Abramson, 1964). On the other hand, the acoustic definition of retroflexion is far more complex. Hamann (2003) mainly focused on F4 and F3 lowering in neighboring vowels transitions (see Fig. 1) and short closure and release phases[17].

[17] As I mentioned above, this property finds a peculiar counterexample in Italian dialects.

Other studies also searched for regularities in spectral moments, i.e. high skewness and kurtosis and low center of gravity and standard deviation. These tendencies depend on the lengthening of the front oral cavity during tongue retraction (see e.g. Tabain, Butcher, Breen & Beare, 2014). In the study previewed here, SOBFFC represents the effect of lateral *continuous* productions on the emergence of retroflex *stops.* For this reason, differently from aspiration, retroflexion will not be treated as a gradient property, but as a class of responses featuring voicing and occlusion. In other words, SOBFFC will not try to predict the degree of retroflexion, but its absolute presence (a "dialectal" response) against its absence (an "Italian" response). While linear mixed effect statistical models can account for continuous response variables (in our case, aspiration), binary responses imply a logistic regression model with mixed effects. Therefore, in the case of retroflexion, SOBFFC will be treated as a fixed effect in a generalized linear mixed model.

Since Pierrehumbert (1994), acceptability judgements found their way into the representation of probabilistic knowledge of sound patterns. However, to date, this methodology has predominantly been used in usage-based syntactic and morpho-syntactic studies, with the underlying assumption that different levels of entrenchment entail proportional perceptions of grammaticality (Bybee, Eddington, 2006: 349). In this framework, researchers questioned the actual correspondence between frequency and acceptability, finding significant discrepancies at the low end of the frequency spectrum (Bermel, Knittl, 2012). Divjak (2017) specifically tried to clarify this issue, finding out that both word frequency and n-gram frequency are worst predictor of acceptability than conditional probabilities, i.e. measures that weight the frequency of occurrence of the inquired element in a specific construction against the overall frequency of the construction (attraction) or of the element (reliance). As this discovery is in line with the FFC tradition, I would argue that an application of these formulae to phonetic acceptability judgments could hold promising results. Research designs interested in observing correspondences between subjects' intuitions and actual use often rely on naturalistic utterances submitted as stimuli (e.g. Bresnan, Ford, 2010). In order to counterbalance potential effects of sentence compositionality beyond the inquired construct, common practices advise for the selection of multiple phrasings containing the test items; as Schütze, Sprouse (2013: 39) advocate eight or more different lexicalizations for syntactic structures, we may suppose that this quantification exponentially grows for small units of analysis, such as specific segments. Indeed, the problem is not posited in inter-categorial analyses, where specific phonetic variants are overall considered predictors of specific judgments against a (e.g., standard) baseline (e.g. Leivada, Papadopoulou, Kambanaros & Grohmann, 2017). However, in our case, the object of interest is intra-categorial, since the posited effect of frequency applies in a gradual manner to the judgements of the same category of items: in other words, the higher the SOBFFC value of a dialect word, the higher this should score in acceptability ratings. The design explicated in Kawahara (2011) could be adapted to this research. Raters provided judgments of naturalness of words containing the inquired phonet-

ic variation with a 5-point scale ranging from very natural to very unnatural. Stimuli were proposed in the template *given [the baseline form], how natural would you find it to pronounce it as [the varied form]?*. Through this phrasing, baseline scores remain implicit; nonetheless, we may assume that competing Italian exemplars are activated and computed against the dialect form, as both are explicitly mentioned in the questionnaire. As shown in Kawahara (2011), participants' ratings can be investigated regressing them against a linear mixed effect model including frequency (in our case, SOBFFC) among its fixed effects.

4. *Conclusions*

In this paper, I outlined the theoretical background of a research in progress in the Italian village of Antona (MS). The research benefits from the acquisitions of usage-based bilingual phonetics: in particular, I tried to adapt the BFFC formula proposed in Brown (2015) to the analysis of dialect-standard advergence. Dialects do not usually have frequency corpora. For this reason, I suggested the integration of subjective frequency estimates into the representation of the competing dialect-standard exemplar aggregations. The selection of geminate lateral retroflexion and voiceless aspiration as target phenomena will set a test bench for both the under-tested assumption that frequency effects account also for non-reducing variation and the plausible application of BFFC to analogical change. For this purpose, I discussed the application of two psycholinguistic methodologies, namely picture naming and acceptability rating, to this specific research, not without proposing solutions to potential discrepancies between common practices and the current design.

Bibliography

Abramowicz, Ł. (2007). Sociolinguistics meets exemplar theory: Frequency and recency effects in (ing). In *University of Pennsylvania Working Papers in Linguistics,* 13, 27-37.

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U.H., & Segui, J. (2004). Predictors of picture naming speed. In *Behavior Research Methods, Instruments, & Computers*, 36(1), 140-155.

Alfonzetti, G. (1992). *Il discorso bilingue. Italiano e dialetto a Catania.* Milano: FrancoAngeli.

Ambrosi, A.C. (1956). Gli attuali limiti dell'area fonetica cacuminale nelle Alpi Apuane. In *Convegno di studi apuani. Rendiconti dei lavori della terza giornata in Carrara – 28 giugno 1955*. Carrara: D. Sanguinetti e figli, 33-40.

Ambrosi, A.C. (1966-1967). Le principali varietà dialettali della Lunigiana. In *La Spezia. Rivista del Comune*, 67, 121-126.

Anderwald, L., Szmrecsanyi, B. (2009). Corpus linguistics and dialectology. In Lüdeling, A., Kytö, M. (Eds.), *Corpus Linguistics. An International Handbook. Vol. II.* Berlin/New York: De Gruyter Mouton, 1126-1140.

Auer, P., Schwarz, C. (2014). Dialect-to-Standard Advergence. The Relevance of Compound Borrowing. In Torres Cacoullos, R., Dion, N. & Lapierre, A. (Eds.), *Linguistic Variation. Confronting Fact and Theory*. London: Routledge, 263-281.

Baayen, R.H. (2001). *Word frequency distributions*. Boston: Kluwer.

Backus, A. (2015). A usage-based approach to code-switching: The need for reconciling structure and function. In Stell, G., Yakpo, K. (Eds.), *Code-switching Between Structural and Sociolinguistic Perspectives.* Berlin/Boston: De Gruyter, 19-38.

Balota, D.A., Pilotti, M., & Cortese, M.J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. In *Memory & Cognition*, 29, 639-647.

Baroni, M., Bernardini, S., Ferraresi, A. & Zanchetta, E. (2009). The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. In *Language Resources and Evaluation*, 43(3), 209-226.

Barras, W., Honeybone, P. & Trousdale, G. (2007). Is word-frequency a factor in phonological demergers? The sociophonology of SQUARE and NURSE in Lancashire English. Paper presented at the *2nd International Conference on the Linguistics of Contemporary English*, Toulouse.

Bermel, N., Knittl, L. (2012). Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech. In *Corpus Linguistics and Linguistic Theory*, 8(2), 241-275.

Berruto, G. (1987). Lingua, dialetto, diglossia, dilalìa. In Holtus, G., Kramer, J. (Eds.), *Romania et Slavia adriatica. Festschrift für Žarko Muljačić*. Hamburg: Buske, 57-81.

Berruto, G. (1989). Tra italiano e dialetto. In Holtus, G., Metzeltin, M. & Pfister, M. (Eds.), *La dialettologia italiana oggi. Studi offerti a Manlio Cortelazzo*. Tübingen: Narr, 107-122.

Berruto, G. (1997a). Code-switching and code-mixing. In Maiden, M., Parry, M. (Eds.), *The Dialects of Italy*. London/ New York: Routledge, 394-400.

Berruto, G. (1997b). Linguistica del contatto e aspetti dell'italianizzazione dei dialetti: appunti di creolistica casalinga. In Holtus, G., Kramer, J. & Schweickard, W. (Eds.), *Italica et Romanica. Festschrift für Max Pfister zum 65. Geburtstag*. Tübingen: Max Niemeyer Verlag, 13-30.

Berruto, G. (2004). Su restrizioni grammaticali nel codemixing e situazioni sociolinguistiche. Annotazioni in margine al modello MLF. In *Sociolinguistica*, 18, 54-72.

Berruto, G. (2005). Dialect/standard convergence, mixing, and models of language contact: the case of Italy. In Auer, P., Hinskens, F. & Kerswill, P. (Eds.), *Dialect Change. Convergence and Divergence in European Languages*. Cambridge: Cambridge University Press, 81-95.

Berruto, G. (2011). Interazione e compenetrabilità di grammatiche nel contatto linguistico. Il caso dell'enunciazione mistilingue. In Bombi, R., D'Agostino, M., Dal Negro, S. & Franceschini, R. (Eds.), *Lingue e culture in contatto. In memoria di Roberto Gusmani. Atti del 10° Congresso dell'Associazione Italiana di Linguistica Applicata*. Perugia: Guerra, 47-70.

Berruto, G. (2018). The languages and dialects of Italy. In Ayres-Bennett, W., Carruthers, J. (Eds.), *Manual of Romance Sociolinguistics*. Berlin: De Gruyter, 494-525.

Bertuccelli, G., Casotti, P. (1982). *Antona. Storia-Leggenda-Arte-Folklore-Architettura-Geografia-Tradizioni-Lavoro-Musica-Dialetto-Medicamenti*. Massa: Assessorato Pubblica Istruzione e Cultura.

Bonin, E. (1952). Beiträge zur Mundart und Volkskunde von Gorfigliano (Garfagnana) und Nachbarorte. Ph.D. Dissertation, University of Munich.

Bottiglioni, G. (1911). Dalla Magra al Frigido. Saggio fonetico. In *Revue de Dialectologie Romane*, 3, 77-143.

Bottiglioni, G. (1926). La penetrazione toscana e le regioni di Pomonte nei parlari di Corsica (Saggio di ricostruzione storico-linguistica). In *L'Italia Dialettale*, 3, 2-69.

Bottiglioni, G. (1930). Di alcune particolari tendenze fonetiche nell'italico (Continuazione e fine). In *Athenaeum. Studii periodici di letteratura e storia dell'antichità*, 8(1), 3-26.

Bresnan, J., Ford, M. (2010). Predicting syntax: Processing dative constructions in American and Australian varieties of English. In *Language*, 86(1), 186-213.

Broersma, M. (2009). Triggered codeswitching between cognate languages. In *Bilingualism: Language and Cognition*, 12(4), 447-462.

Broersma, M., Carter, D. & Acheson, D.J. (2016). Cognate Costs in Bilingual Speech Production: Evidence from Language Switching. In *frontiers in Psychology*, 7, 1-16.

Broersma, M., Carter, D., Donnelly, K. & Konopka, A (2019). Triggered codeswitching: Lexical processing and conversational dynamics. In *Bilingualism: Language and Cognition*, ahead of print, 1-14.

Broersma, M., De Bot, K. (2006). Triggered codeswitching: A corpus-based evaluation of the original triggering hypothesis and a new alternative. In *Bilingualism: Language and Cognition*, 9(1), 1-13.

Brown, E.L. (2004). Reduction of syllable-initial /s/ in the Spanish of New Mexico and Southern Colorado: A usage-based approach. Ph.D. Dissertation, University of New Mexico.

Brown, E.L. (2013). Word classes in studies of phonological variation: Conditioning factors or epiphenomena?. In Howe, C., Blackwell, S. & Lubbers Quesada, M. (Eds.), *Selected Proceedings of the 15th Hispanic Linguistics Symposium*. Somerville: Cascadilla Proceedings Project, 179-186.

Brown, E.L. (2015). The role of discourse context frequency in phonological variation: A usage-based approach to bilingual speech production. In *International Journal of Bilingualism*, 19(4), 387-406.

Brown, E.L. (2018). Cumulative exposure to phonetic reducing environments marks the lexicon: Spanish /d-/ words spoken in isolation. In Smith, K.A., Nordquist, D. (Eds.), *Functionalist and Usage-based Approaches to the Study of Language: In honor of Joan L. Bybee*. Amsterdam: John Benjamins, 127-153.

Brown, E.L., Amengual, M. (2015). Fine-grained and probabilistic cross-linguistic influence in the pronunciation of cognates: Evidence from corpus-based spontaneous conversation and experimentally elicited data. In *Studies in Hispanic and Lusophone Linguistics*, 8(1), 59-83.

Brown, E.L., Harper, D. (2009). Phonological Evidence of Interlingual Exemplar Connections. In *Studies in Hispanic and Lusophone Linguistics*, 2(2), 257-274.

Brown, E.L., Raymond, W.D. (2012). Explaining the distribution of Spanish f-/h- words. In *Diachronica*, 29(2), 139-161.

Bybee, J.L. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, J.L. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. In *Language Variation and Change*, 14, 261-290.

Bybee, J.L. (2006). From usage to grammar. The mind's response to repetition. In *Language*, 82, 711-733.

Bybee, J.L., Beckner, C. (2009). Usage-based theory. In Heine, B., Narrog, H. (Eds), *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, 827-855.

Bybee, J.L., Eddington, D. (2006). A Usage-based Approach to Spanish Verbs of Becoming. In *Language*, 82(2), 323-355.

Calamai, S. (2017). Tuscan between standard and vernacular: a sociophonetic perspective. In Cerruti, M., Crocco, C. & Marzo, S. (Eds.), *Towards a New Standard. Theoretical and Empirical Studies on the Restandardization of Italian*. Boston: De Gruyter, 213-241.

Carpitelli, E. (1995). Description des systèmes des voyelles toniques de quelques dialectes de la Toscane nord-occidentale. In *Géolinguistique*, 6, 43-73.

Carpitelli, E. (2017). Esempi di variazione fonetica nell'Appennino settentrionale. Il caso dei vocalismi instabili. In Regueira, X.L., Fernández Rei, E. (Eds.), *Estudos sobre o cambio lingüístico no galego actual*. Santiago de Compostela: Consello da Cultura Galega, 215-240.

Carroll, J.B. (1971). Measurement properties of subjective magnitude estimates of word frequency. In *Journal of Verbal Learning and Verbal Behavior*, 10, 722-729.

Celata, C. (2006). Analisi dei processi di retroflessione delle liquide in area romanza con dati sperimentali dal còrso e dal siciliano. Ph.D. Dissertation, Scuola Normale Superiore di Pisa.

Cenceschi, S. Trivilini, A. Sbattella, L. & Tedesco, R. (2019). Collecting Italian spontaneous social media speech: the WAsp2 project. Poster presented at the *XV AISV Conference (Audio Archives at the Crossroads of Speech Sciences, Digital Humanities and Digital Heritage)*, Arezzo, February 14-16 2019.

Cerruti, M. (2016). L'italianizzazione dei dialetti: una rassegna. In *Quaderns d'Italià*, 21, 63-74.

Cerruti, M. (2018). Code-switching in Italo-Romance: a variationist study of convergence in bilingual speech. In *Lingue e Linguaggio*, 1/2018, 87-106.

Cerruti, M., Regis, R. (2005). 'Code switching' e teoria linguistica: la situazione italoromanza. In *Italian Journal of Linguistics*. *Rivista di Linguistica*, 17(1), 179-208.

Cho, T., Ladefoged, P. (1999). Variation and Universals in VOT: Evidence from 18 Languages. In *Journal of Phonetics*, 27, 207-229.

Clark, L. (2008). Re-examining vocalic variation in Scottish English: a Cognitive Grammar approach. In *Language Variation and Change*, 20(3), 255-273.

Clark, L., Watson, K. (2011). Testing claims of a usage-based phonology with Liverpool English t-to-r. In *English Language and Linguistics*, 15(3), 523-547.

Clopper, C.G., Mitsch, J.F. & Tamati, T.N. (2017). Effects of phonetic reduction and dialect variation on vowel production. In *Journal of Phonetics*, 60, 38-59.

Clyne, M. (1967). *Transference and triggering: Observations on the language assimilation of postwar German-speaking migrants in Australia.* The Hague: Martinus Nijhoff.

Coetzee, A.W., Kawahara, S. (2013). Frequency biases in phonological variation. In *Natural Language and Linguistic Theory*, 30(1), 47-89.

Coseriu, E. (1980)."Historische sprache" und "dialekt". In Göschel, J., Ivić, P. & Kehr, K. (Eds.), *Dialekt und dialektologie. Ergebnisse des Internationalen Symposions „Zur Theorie des Dialekts", Marburg/Lahn 5. - 10. September 1977.* Wiesbaden: Steiner, 106-122.

Dal Negro, S., Vietti, A. (2011). Italian and Italo-Romance Dialects. In *International Journal of the Sociology of Language*, 210, 71-92.

De Schryver, J., Neijt, A., Ghesquière, P. & Ernestus, M. (2008). Analogy, Frequency, and Sound Change. The Case of Dutch Devoicing. In *Journal of Germanic Linguistics*, 20(2), 159-195.

Della Sala, A.M. (1986). Caratteristiche morfologiche e sociali del linguaggio di Antona. In *Le Apuane. Rivista di cultura-storia-etnologia*, 11, 37-56.

Dinkin, A.J. (2008). The real effect of word frequency on phonetic variation. In *University of Pennsylvania Working Papers in Linguistics*, 14, 97-106.

Divjak, D. (2017). The Role of Lexical Frequency in the Acceptability of Syntactic Variants: Evidence From *that*-Clauses in Polish. In *Cognitive Science. A Multidisciplinary Journal*, 41(2), 354-382.

Dollinger, S. (2015). *The Written Questionnaire in Social Dialectology, History, theory, practice.* Amsterdam/Philadelphia: John Benjamins.

Drager, K. (2011). Sociophonetic variation and the lemma. In *Journal of Phonetics*, 39, 694-707.

Eddington, D., Channer, C. (2010). American English Has Go? A Lo? Of Glottal Stops: Social Diffusion and Linguistic Motivation. In *American Speech*, 85(3), 338-351.

Forrest, J. (2017). The dynamic interaction between lexical and contextual frequency: A case study of (ING). In *Language Variation and Change*, 29, 129-156.

Foulkes, P. (2010). Exploring social-indexical knowledge: A long past but a short history. In *Laboratory Phonology*, 1(1), 5-39.

Ghimenton, A. (2013a). Quantitative Approaches to Italian / Dialect Family Interactions: Considerations of Methodology and Language Transmission in a Contact Situation. In *Journal of Language Contact*, 6, 106-133.

Ghimenton, A. (2013b). Pathways to multilingual acquisition in Veneto: a usage-based perspective of code choices in the input and output in a language contact situation. In Reif, M., Robinson, J. & Pütz, M. (Eds.), *Variation in Language and Language Use. Linguistic, Socio-Cultural and Cognitive Perspectives*. Bern: Peter Lang, 59-72.

Giannarelli, D. (1912). *Caratteri generali dei dialetti lunigianesi compresi fra la Magra e l'Appennino Reggiano.* Tortona: Tipografia Peila.

Giannelli, L. (2000). *Toscana*. Pisa: Pacini.

Giannelli, L., Savoia, L.M. (1979-1980). L'indebolimento consonantico in Toscana. In *Rivista Italiana di Dialettologia. Scuola società territorio*, 3-4, 38-101.

Habib, R. (2011). Frequency effects and lexical split in the use of [t] and [s] and [d] and [z] in the Syrian Arabic of Christian rural migrants. In *Journal of Historical Linguistics*, 1(1), 77-105.

Hamann, S.R. (2003). *The Phonetics and Phonology of Retroflexes.* Utrecht: LOT.

Hay, J., Foulkes, P. (2016). The evolution of medial /t/ over real and remembered time. In *Language*, 92(2), 298-330.

Hay, J., Jannedy, S., & Mendoza-Denton, N. (1999). Oprah and /ay/: Lexical frequency, referee design, and style. In *Proceedings of the 14th international congress of phonetic sciences*, 1389-1392.

Hoshino, N., Kroll, J.F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script?. In *Cognition*, 106, 501-511.

Howes, D. (1954). On the interpretation of word frequency as a variable affecting speed of recognition. In *Journal of Experimental Psychology*, 48(2), 106-112.

ISTAT (2017). *L'uso della lingua italiana, dei dialetti e delle lingue straniere (anno 2015).* National survey.

Johnson, K. (1997). Speech perception without speaker normalization. An exemplar model. In Johnson, K., Mullennix, J.W. (Eds.), *Talker variability in speech processing*. San Diego: Academic Press, 145-165.

Kawahara, S. (2011). Aspects of Japanese loanword devoicing. In *Journal of East Asian Linguistics*, 20, 169-194.

Langacker, R. (1987). *Foundations of cognitive grammar. Vol. 1. Theoretical prerequisites.* Stanford: Stanford University Press.

Lee, O., Kapatsinski, V. (2014). Frequency Effects in Morphologisation of Korean /n/-Epenthesis. Paper presented as a poster at the *89th annual meeting of the Linguistic Society of America* (Portland, Oregon, January 2015).

Leivada, E., Papadopoulou, E., Kambanaros, M. & Grohmann, K.K. (2017). The Influence of Bilectalism and Non-standardization on the Perception of Native Grammatical Variants. In *frontiers in Psychology*, 8, 1-11.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, W.J., Marchal, A. (Eds.), *Speech production and speech modelling*. Dordrecht: Kluwer Academic Publishers, 403-439.

Lisker, L., Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. In *Word*, 20(3), 527-565.

Loporcaro, M. (1997). Lengthening and raddoppiamento fonosintattico. In Maiden, M., Parry, M. (Eds.), *The Dialects of Italy*. London/New York: Routledge, 41-51.

Maffei Bellucci, P. (1977). *Lunigiana*. Pisa: Pacini.

Maiden, M., Parry, M. (Eds., 1997). *The Dialects of Italy*. London/New York: Routledge.

Mariotti Giromella, N. (2006). Preghiere in dialetto antonese. In *Le Apuane. Rivista di cultura-storia-etnologia*, 51, 140-144.

Mattheier, K.J. (1996). Varietätenkonvergenz: Überlegungen zu einem Baustein einer Theorie der Sprachvariation. In *Sociolinguistica*, 10, 31-52.

Mello, E., Miozzo, M. & Peressotti, F. (2017). Decision making in a not-so-foreign language: is a there a reduction of decision biases in Venetian dialect? Poster presented at the *XX European Society for Cognitive Psychology Conference (ESCOP)*, Postdam, Germany.

Merlo, C. (1956). Liguri o mediterranei gli odierni Apuani?. In *L'Italia Dialettale*, 20, 86.

Mioni, A.M. (1992). Il punto di vista della sociolinguistica. In Foresti, F., Sobrero, A.A. (Eds.), *Dialetti realtà ricerca. Atti del colloquio "I dialetti e la dialettologia negli anni Novanta"*. Bologna: CLUEB, 17-34.

Muldner, K., Hoiting, L., Sanger, L., Blumenfeld, L. & Toivonen, I. (2019). The phonetics of code-switched vowels. In *International Journal of Bilingualism*, 23(1), 37-52.

Myers, J., Li, Y. (2009). Lexical frequency effects in Taiwan Southern Min syllable contraction. In *Journal of Phonetics*, 37, 212-230.

Nadeu, M. (2016). Phonetic and phonological vowel reduction in Central Catalan. In *Journal of the International Phonetic Association*, 46(1), 33-60.

Navarrete, E., Arcara, G., Mondini, S. & Penolazzi, B. (2019). Italian norms and naming latencies for 357 high quality color images. In *PLoS ONE*, 14(2), 1-17.

Nusbaum, H.C., Pisoni, D.B. & Davis, C.K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. In *Research in Speech Perception. Progress Report. Indiana University*, 10, 357-375.

Pellegrini, G.B. (1980^2). *Carta dei Dialetti d'Italia*. Pisa: Pacini.

Phillips, B. (2006). *Word frequency and lexical diffusion*. New York: Palgrave.

Pierrehumbert, J.B. (1994). Syllable structure and word structure: a study of triconsonantal clusters in English. In Keating, P.A. (Ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*. Cambridge: Cambridge University Press, 168-188.

Pierrehumbert, J.B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In Bybee, J., Hopper, P. (Eds), *Frequency effects and the emergence of lexical structure*. Amsterdam: John Benjamins, 137-157.

Pierrehumbert, J.B. (2002). Word-specific phonetics. In Gussenhoven, C., Warner, N. (Eds.), *Laboratory phonology VII*. Berlin: Mouton de Gruyter, 101-140.

Raymond, W.D., Brown, E.L. (2012). Are effects of word frequency effects of context of use? An analysis of initial fricative reduction in Spanish. In Gries, S., Divjak, D. (Eds.), *Frequency effects in language. Vol. 2. Learning and processing*. Berlin: Mouton de Gruyter, 35-52.

Raymond, W.D., Brown, E.L., & Healy, A.F. (2016). Cumulative context effects and variant lexical representations: Word use and English final t/d deletion. In *Language Variation and Change*, 28, 175-202.

Regis, R. (2013). Può un dialetto essere standard?. In *Vox Romanica*, 72, 151-169.

Regis, R. (2017). La nozione coseriana di dialetto e le sue implicazioni per l'area italoromanza. In *Revue Romane. Langue et littérature. International Journal of Romance Languages and Literatures*, 52(2), 137-169.

Ringeling, J.C.T. (1984). Subjective Estimations as a Useful Alternative to Word Frequency Counts. In *Interlanguage Studies Bulletin*, 8, 59-69.

Rossi, M. (1976). *Contribution à la méthodologie de l'analyse linguistique avec application à la description phonétique et phonologique du parler de Rossano (province de Massa, Italie): thèse*

presentée devant l'Université de Paris III le 10 juin 1974. Lille/Paris: Atelier Reproduction des thèses / Librairie Honoré Champion.

Sanga, G. (1985). La convergenza linguistica. (Alla memoria di Benvenuto Terracini). In *Rivista Italiana di Dialettologia. Società scuola territorio*, 9, 7-41.

Savoia, L.M. (1980). Fonologia delle varietà apuane e garfagnine: consonantismo. In *Studi Urbinati di Storia, Filosofia e Letteratura*, 2, 233-293.

Schmid, H.J. (2000). *English Abstract Nouns as Conceptual Shells. From corpus to cognition.* Berlin/New York: Mouton de Gruyter.

Schmitt, J.M., Auer, P. & Ferstl, E.C. (2019). Understanding fairy tales spoken in dialect: an fMRI study. In *Language, Cognition and Neuroscience*, 34(4), 440-456.

Schütze, C., Sprouse, J. (2013). Judgment data. In Podesva, R.J., Sharma, D. (Eds.), *Research methods in linguistics*. Cambridge: Cambridge University Press, 27-50.

Sobrero, A.A. (1996). Italianization and Variations in the Repertoire: The Koinai. In *Sociolinguistica. International Yearbook of European Sociolinguistics*, 10, 105-111.

Song, J., Dalola, A. (2019). Linguistic entrenchment and the effect of subjective lexical familiarity in Korean /n/-insertion. In *Proceedings of the Linguistic Society of America*, 4(34), 1-8.

Sorianello, P. (1996). Indici fonetici delle occlusive sorde in cosentino. In *Rivista Italiana di Dialettologia*, *Lingua dialetti società*, 20, 123-159.

Tabain, M., Butcher, A., Breen, G. & Beare, R. (2014). A spectral analysis of laterals in three Central Australian languages. In Hay, J., Parnell, E. (Eds.), *Proceedings of the 15th Australasian Speech Science and Technology Association*, 98-101.

Tagliamonte, S.A. (2013). Comparative Sociolinguistics. In Chambers, J.K., Schilling, N. (Eds.), *The Handbook of Language Variation and Change. Second Edition*. Boston: John Wiley & Sons, 128-156.

Thompson, G.L., Desrochers, A. (2009). Corroborating biased indicators: Global and local agreement among objective and subjective estimates of printed word frequency. In *Behavior Research Methods*, 41(2), 452-471.

Thorndike, E.G., Lorge, L. (1944). *A teacher's word book of 30,000 words*. New York: Columbia University Press.

Todd, S., Pierrehumbert, J.B. & Hay, J. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. In *Cognition*, 185, 1-20.

Toglia, M.P., Battig, W.F. (1978). *Handbook of semantic word norms*. Hillsdale: Erlbaum.

Trifone, P. (1992). *Roma e il Lazio*. Torino: UTET.

Wieling, M., Montemagni, S., Nerbonne, J. & Baayen, H. (2014). Lexical differences between Tuscan dialects and Standard Italian: A sociolinguistic analysis using generalized additive mixed modeling. In *Language*, 90(3), 669-692.

Yan, H. (2018). *The Nature of Variation in Tone Sandhi Patterns of Shanghai and Wuxi Wu.* Singapore: Springer/Peking University Press.

Yan, H., Zhang, J. (2017). The nature of variation in the tone Sandhi patterns of Wuxi Wu. In *The Journal of the Acoustical Society of America*, 142, 2519.

Zhang, J., Meng, Y. (2016). Structure-dependent tone sandhi in real and nonce disyllables in Shanghai Wu. In *Journal of Phonetics*, 54, 169-201.