languages

MDPI

Article

Lexical Frequency and the Realization of Italian Dental Affricates

Chiara Meluzzi * and Nicholas Nese

Department of Literary, Philological and Linguistic Studies, University of Milan, 20122 Milano, Italy; nicholas.nese@unimi.it

* Correspondence: chiara.meluzzi@unimi.it

Abstract

This study investigates the phonetic variation in Italian dental affricates, focusing on the role of lexical frequency, phonological context, geographical origin, and speech style (read speech vs. collaborative dialogue). Previous research on lexically based phonetics has emphasised the link between item frequency and phonetic realisation; in this paper, we test these premises on a class of rare and marked sounds—dental affricates. A corpus of read sentences and map-task dialogues produced by northern and southern Italian speakers was analysed acoustically with respect to voicing and duration. Results show that phonological context and geographical origin are the primary determinants of voicing, with southern speakers favouring voiced realisations and word-initial position strongly conditioning voicing patterns. Lexical frequency does not significantly predict voicing category once between-speaker and between-item variability are appropriately modelled, but it does exert a positive effect on affricate duration: higher-frequency words contain relatively longer affricates, reflecting compensatory preservation of segmental identity within otherwise reduced words. Speech style significantly affects duration, with reading favouring longer realisations. These findings reveal a dissociation between categorical and gradient levels of phonetic variation, supporting usage-based models in which lexical frequency modulates fine-grained phonetic implementation rather than determining phonological categorisation.

Keywords: Italian phonetics; dental affricates; lexical frequency; phonetic variation; regional variation; speech style

1. Introduction

Recent advances in phonetic research (including the ones collected in this Special Issue for Italian) have demonstrated that fine-grained acoustic variation systematically encodes lexical and grammatical information beyond traditional phonological contrasts. Converging evidence reveals that high-level linguistic properties, including lexical frequency, significantly influence phonetic realisation, with listeners demonstrating sensitivity to these subtle acoustic cues during word recognition and processing (Gahl, 2008; Gahl & Strand, 2016). Particularly compelling evidence comes from studies showing that distinct homophones exhibit systematic acoustic differences, suggesting that speakers encode lexical information through phonetic detail that listeners can reliably decode (Martinuzzi & Schertz, 2022). These frequency-driven effects on phonetic realisation have been documented across multiple languages and phonetic parameters, from segmental duration (Pluymaekers et al., 2005) to suprasegmental features such as lexical tone (Bi & Chen, 2022).

The present study contributes to this growing literature by investigating how lexical frequency modulates the phonetic realisation of Italian dental affricates. Indeed, Italian

Academic Editor: Chiara Celata

Received: 22 October 2024
Revised: 2 April 2026
Accepted: 7 April 2026
Published: 1 May 2026

is one of the few languages in Europe (and the only one among the Romance group) to present a phonological difference between voiced and voiceless dental affricates. Nevertheless, dental affricates occur with low frequency in modern Italian vocabulary, and the words that contain them display heterogeneous etymological origins, reflecting diverse historical and linguistic influences on the Italian lexicon. Thus, the voicing phonological contrast in Italian dental affricates could offer unique insights into the interaction between statistical word properties and articulatory implementation in a language with systematic gemination contrasts.

In this work, the influence of lexical frequency and etymology in Italian dental affricates will be tested in two varieties of regional Italian (i.e., North-Western and Sicilian). Although traditionally Italian phonology labels these sounds as "alveolar", scholars working on the topic have variously emphasised the variability in possible places of articulation of these sounds in correlation with either geographic variability (Meluzzi, 2020) or stylistic variability (Nese, 2023). Therefore, the label "dental" is here preferred as a broad category, in line with previous works on the topic.

The paper is structured as follows: Section 2 reviews relevant literature on lexical frequency effects and Italian dental affricates. Section 3 outlines our research questions, data collection, and annotation methods. Section 4 presents statistical analyses of voicing and duration patterns. Section 5 discusses the findings, and Section 6 concludes.

2. Theoretical Remarks

This section reviews two areas essential to the present investigation: the role of lexical frequency in shaping phonetic variation (Section 2.1) and the phonetic and sociolinguistic characteristics of Italian dental affricates (Section 2.2).

2.1. Phonetic Variation and the Lexicon

A growing amount of phonetic research has provided empirical evidence demonstrating that fine-grained phonetic details carry significant linguistic information and systematically vary according to a variety of lexical, grammatical, and contextual factors (Pierrehumbert, 2002; Johnson, 2006), other than social and individual ones. Lexical frequency represents one of the most robust predictors of phonetic variation, with high-frequency words systematically showing greater acoustic reduction compared to low-frequency words across multiple acoustic dimensions (Jurafsky et al., 2001; Bell et al., 2009). This phenomenon, known as frequency-based reduction, has been documented in numerous languages and is typically manifested through shortened duration, centralised vowels, lenited consonants, and reduced articulatory precision in frequent words (Pluymaekers et al., 2005; Gahl, 2008).

The theoretical foundation for frequency effects on phonetic implementation rests on the principle of communicative efficiency, whereby speakers optimise their articulatory effort according to the predictability and accessibility of lexical items (Aylett & Turk, 2004; Munson & Solomon, 2004). It is commonly stated that high-frequency words are more predictable and more easily accessed from the mental lexicon, allowing speakers to reduce articulatory effort while maintaining communicative effectiveness. Conversely, low-frequency words require greater articulatory precision to ensure successful recognition by listeners.

Different explanations compete to explain this pattern. The usage-based model suggests that repeated production of high-frequency words leads to routinization and articulatory automation, resulting in progressive reduction over time (Bybee, 2001; Phillips, 2006). This view is also supported by evidence from sound changes in progress, where high-frequency words often lead phonetic innovations that subsequently spread to lower-

frequency items. The exemplar-based approach proposes that lexical representations consist of clouds of phonetic exemplars stored in memory, with high-frequency words characterised by denser exemplar clouds that include more reduced variants (Johnson, 1997; Pierrehumbert, 2002). During production, speakers select from this range of stored exemplars, with the probability of producing reduced variants increasing with lexical frequency. Alternative accounts emphasise the role of predictability and informativity in shaping phonetic implementation. According to the Smooth Signal Redundancy Hypothesis (Aylett & Turk, 2004), speakers modulate articulatory effort inversely to the predictability of linguistic elements. High-frequency words are more predictable in context and therefore require less acoustic energy to be successfully recognised.

Frequency effects on phonetic implementation have been documented across typologically diverse languages, suggesting universal principles underlying this phenomenon. In English, numerous studies have demonstrated duration reduction in high-frequency words (Bell et al., 2009; Gahl, 2008), with additional effects on vowel quality (Munson & Solomon, 2004) and consonant realisation (Raymond et al., 2006). Similar patterns have been observed in other Germanic languages, including Dutch (Pluymaekers et al., 2005) and German (Schmitz & Baer-Henney, 2024), as well as in Romance languages such as Spanish (Torreira & Ernestus, 2011) and French (Adda-Decker et al., 2005). Cross-linguistic consistency in frequency effects suggests that these phenomena reflect fundamental properties of speech production and lexical access rather than language-specific organisational principles. Italian has received increasing attention in research on fine-grained phonetic variation, with studies documenting systematic acoustic differences related to linguistic (and also extra-linguistic) factors. Recent experimental work on different languages has documented systematic phonetic variation conditioned by lexical and morphological properties. For instance, recent studies in Italian examining morphologically complex words have revealed that speakers systematically encode lexical information through fine-grained acoustic variation (Rossi et al., forthcoming), with effects modulated by factors including transparency, productivity, and relative frequency of word forms and their bases. Lexical frequency has been shown to significantly influence phonetic realisation, with the most robust effects documented for segmental duration, though additional effects have been observed across various acoustic parameters. For instance, high-frequency words in Italian show duration reduction (Savy, 1999) compared to low-frequency words. These effects interact with factors including syllable structure, stress position, and regional variety, creating complex patterns of phonetic variation within the Italian lexicon (see also Schettino & Cotugno, 2025). Moreover, the effect of lexical frequency has also been addressed in Italian L2 from a psycholinguistic perspective (e.g., Bellocchi et al., 2016), and in words vs. non-words production (Paizi et al., 2010), thus confirming the role of frequency in shaping phonetic production. However, French data in Meunier and Espesser (2011) have found no direct evidence of lexical frequency on vowel reduction, in terms of both duration and centralization. Their results show how vowels in final syllables tend to be reduced in monosyllabic function words compared to monosyllabic content words, thus emphasising the intertwined role of phonetics, morphology, and the lexicon.

Crucially, research on frequency-related phonetic reduction has increasingly extended beyond European languages, revealing both universal patterns and language-specific modulations that also align within an Exemplar Theory framework. In Japanese, Hashimoto (2021) demonstrated that morpheme duration is systematically affected by three types of information-theoretic measures: morpheme frequency, contextual morpheme predictability (both forward and backward), and average morpheme predictability (henceforth, informativity). His findings show that morphemes with higher frequency and higher contextual predictability are produced with shorter duration; backward predictability (conditioned on

the following morpheme) exerts stronger effects than forward predictability. A significant effect of informativity was also found, thus suggesting that reduction patterns may be lexicalized rather than purely context-dependent. In a following (and complementary) study focusing on verbal conjugation, Hashimoto (2023) found that Japanese non-past indicative forms with higher conjugation predictability (i.e., verbs frequently used in that particular form) were produced with shorter duration, with an estimated reduction of 7.7 milliseconds when comparing mean conjugation predictability to one standard deviation lower. These findings were interpreted within Exemplar Theory as reflecting ease of production target creation: conjugation forms with higher resting activation levels (due to more frequent usage) are accessed more quickly, resulting in reduced articulatory effort and shorter duration.

Similar patterns have been documented in other tone languages. Data on Dalian Mandarin in Bi and Chen (2022) have shown how tonal neutralisation is lightly predicted by lexical frequency per se, whereas homophone neighbourhood density significantly affects both tonal realisation and the maintenance of tonal contrasts, with low-density syllables preserving more acoustic differentiation between the two falling tones than high-density syllables. In Taiwan Southern Min, Wang (2022) investigated how predictability measurements-including bigram surprisal, bigram informativity, and lexical frequency-interact with prosodic phrasing to affect syllable duration in spontaneous speech. The study found that higher informativity and surprisal led to longer syllables, consistent with information-theoretic predictions. Importantly, Wang demonstrated that these predictability effects were modulated by prosodic position: there was a general weakening of predictability effects for syllables closer to prosodic boundaries, especially in pre-boundary positions where pre-boundary lengthening was strongest. However, the effect of word informativity appeared least modulated by boundary marking, suggesting that informativity-specific durational variants may be stored as part of lexical representations. In Mandarin, Tang and Shaw (2021) provided converging evidence that prosodic information "leaks into" the mental representations of words, with word-specific duration patterns reflecting both local contextual predictability and prosodic characteristics of typical usage contexts.

When considered alongside findings from European languages, cross-linguistic evidence reinforces the central role of frequency in shaping phonetic realisation. The finding that words with generally high predictability show reduction even in locally unpredictable contexts (documented in Japanese, Mandarin, and Taiwan Southern Min) suggests that phonetic variants are stored as part of lexical representations rather than computed purely online. This lexicalization hypothesis receives further support from the observation that informativity effects are more resistant to modulation by prosodic context than contextual surprisal effects, as demonstrated particularly clearly in Wang's (2022) analysis of Taiwan Southern Min. The consistency of these patterns across languages with radically different prosodic systems (i.e., stress-timing, syllable-timing, and mora-timing) seems to indicate that the storage of usage-based phonetic variants may be a universal property of the human lexical system.

2.2. Italian Phonology and Dental Affricates Variability

Among the languages of Europe, only Italian and Polish maintain a phonological opposition between voiceless and voiced dental affricates, whereas other languages present only the voiceless phoneme (see also Meluzzi, 2021, for a diachronic account of dental affricates in Romance languages). In various studies, Marzena Żygis has argued that the voiced dental affricates are prone to phonetic reduction and loss both interlinguistically and diachronically because of their articulatory and aerodynamic properties (cf. Żygis,

2008, but also the remarks in Solé, 2015). Indeed, these properties were at the base of the evolution of affricates in Romance languages, with a general tendency towards a reduction into fricatives and, secondly, a palatalization (Meluzzi, 2021, p. 110). At present, only Catalan and Italian present a voiced–voiceless opposition, but only for Italian scholars generally agree that this opposition also has a phonological (i.e., distinctive) value (see Recasens & Espinosa, 2007; Wheeler, 2005 for Catalan dialectal variation; De Dominicis, 1999, for Italian phonological repertoire).

However, it should be noted that Italian dental affricates present only a few real minimal pairs (De Dominicis, 1999), but only one is really attested, with the opposition between ['rat.tsa] "race" and ['rad.dza] "stingray". This is also due to the different origins of the two Italian dental affricates, whose distribution is historically linked to the evolution of the Italian lexicon in contact with different languages. As summarised in Celata (2004), and later in Meluzzi (2021, pp. 39–40), dental affricates are both the result of an evolution of Latin -TJ- and -DJ-, in which the second element, most likely a palatal approximant, started to be realised as fricative at the beginning of the 2nd c. A.D., and two centuries later, some of these realisations would have been further palatalized by maintaining an affricate manner of articulation. Nowadays, Italian dental affricates derived from Latin roots are mainly found in intervocalic positions, whereas affricates at the beginning of a word are most likely derived from loanwords, both from Germanic languages (e.g., *zaino* "rucksack") or Arabic, through Spanish (e.g., *zucchero* "sugar").

If diachronic changes and contacts may explain the different distribution of Italian dental affricates within the lexicon, it is still not clear which factors correlate with their geographic variability within regional varieties. This variability has long been discussed by dialectologists (e.g., Canepari, 1980; Telmon, 2003, only to name a few), who have highlighted two main sources of variability: the distribution of voicing across phonological contexts and the place of articulation. For the latter point, Canepari (1980) emphasises that dental affricates could be realised as inter-dental or with a "strong shrinking" of the oral cavity during the production of the fricative element; although it is not completely transparent what the author meant with "strong shrinking", the term seems to indicate an articulatory reduction leading to a realisation that Italian phoneticians classified as "solcata" (engl. grooved, cf. Ladefoged & Maddieson, 1996). These realisations are more frequently realised in the Venetian and Trentino regional varieties of Italian. As for voicing, there is no clear-cut distribution between northern and southern varieties of Italian, although in southern varieties, a certain preference for the realisation as voiced in post-consonant position and, in particular, post-lateral ones, such as in *alzare* "to lift", is attested, realised mainly as [al.'dza:.ɾe] (cf. Meluzzi, 2020, p. 43, and the dialectological literature quoted therein). However, previous works have emphasised how dental affricates could be realised as voiced or voiceless by the same speaker in the same phonological context: this intra-speaker variability is only partially explained by the correlation between dental affricate voicing and lexical etymology, but it appears to be linked to stylistic factors (see, for instance, the results in Nese & Meluzzi, 2017; Sbacco & Meluzzi, 2023).

Furthermore, an intermediate voicing degree has been found through acoustic analysis of dental affricates: these intermediate realisations always present a first occlusive segment realised as voiced and a following devoiced (or voiceless) fricative segment. The duration of these variants is also intermediate between voiced and voiceless affricates, albeit more similar to the voiced ones (Meluzzi, 2016). Previous works on different speech communities have shown the emergence of these intermediate realisations in contact situations, in particular in the case of prolonged contact between speakers with a different distribution of voicing of dental affricates in their regional varieties (cf. Nese & Meluzzi, 2017; Meluzzi, 2020; Sbacco & Meluzzi, 2023). The intermediate affricate variant is also

prone to sociophonetic variation since it is strongly influenced by the speaker's sex and age as well as the amount of contact with different Italian regional varieties (see Meluzzi, 2016, for a detailed account focusing on the multilingual town of Bozen, South Tyrol).

As for durational cues, manuals of Italian phonetics and phonology signal how dental affricates are among the five sounds always realised as long when in the intervocalic position. This means that, albeit a graphemic difference, there is no variation in the duration of [ts] in the words *dazi* 'tolls' and *pazzi* 'crazy (pl.)'. A previous sociophonetic investigation on a corpus of 42 Italian speakers has confirmed that intervocalic dental affricates are always realised as long, also with a shortening effect on the previous vowel (but see Mairano et al., 2025, on a possible effect of writing on durational properties of dental affricates). However, data have also highlighted the presence of phonetic shortening of the dental affricate and, in particular, of the plosive phase when they occur in post-sonorant contexts and, in particular, in post-nasal position (e.g., *pinza* 'pliers', cf. Meluzzi, 2016, 2020).

3. Methods and Materials

3.1. Research Questions

Based on previous literature on the interface of phonetics and the lexicon (Section 2.1), in this work, we aim to understand the influence played by lexical factors such as word frequency and etymology in shaping the phonetic variability of dental affricates in two speech styles, namely dialogues vs. read speech. As previous experimental data have shown (cf. Section 2.2), dental affricates in Italian have different historical origins, and their overall frequency is quite limited in the lexicon. This means that, despite some highly frequent words containing a dental affricate (one for all the frequent and typically Italian swearwords, *cazzo* 'dick/fuck'), others may be less represented in the contemporary Italian lexicon and/or belong to less frequent or ancient words. This undoubtedly constitutes a limit for the research design, as we will explain later (cf. Section 3.2), but it is an inherent feature of the Italian lexical–phonological interface.

Specifically, we set an experiment to answer the following research questions:

(1) Does lexical frequency predict systematic differences in affricate realisation?
(2) How does speech style (read vs. dialogue) shape affricate production, and how does this interact with lexical frequency?
(3) Do other sociolinguistic factors contribute to explaining dental affricate variability?

Our hypothesis is that rare and old words will show a major degree of variability, in particular for durational cues, and tend to vary more across speech styles than more lexically frequent and known words.

3.2. Research Design

The study involved eight native Italian university students from Lombardy and Sicily, with four males and four females taking part in the experiment, equally distributed between northern (i.e., Lombard) and southern (i.e., Sicilian) speakers. This means that, for the present study, we included two female speakers and two male speakers from Lombardy and two male and two female speakers from Sicily. Participants were aged between 19 and 26, and they did not recall any speech or hearing disorders. They all signed an informed consent form and agreed to voluntary participation in the experiment.

The experimental design aimed at eliciting both controlled and spontaneous speech reading, since previous works have emphasised durational and voicing differences in dental affricate realisation across speech styles. Indeed, Nese (2023) highlighted a tendency towards voiceless realisations [ts] in read speech in all phonological contexts but the initial one, where the voiced affricate [dz] prevails regardless of the speaker's origin. It was argued

that the voiceless realisation is perceived as more prestigious, especially in post-sonorant contexts, due to its association with northern speech patterns.

The reading task consisted of 48 sentences of equal length and controlled prosodic contour, with 51 target words, since 3 sentences contained two target words (see Supplementary Materials). A standard framework like 'Say X again' was discarded in favour of more realistic sentences (e.g., *Via Mazzini è senza luce*, literally "Mazzini street has no light", i.e., "There's a power outage on Mazzini street"). A complete list of target words and read sentences is provided in the Supplementary Materials. It is important to emphasise that the 48 target words were divided to cover the different phonological contexts and mirror the distribution of these sounds in the Italian lexicon; thus, 10 target words were introduced in initial position (e.g., *zampa* "pow"), 25 in intervocalic position (either spelled as singleton, e.g., *negozi* "shops" or geminate, e.g., *piazza* "square"), 16 in post-voiced position /n/, /r/ and /l/ (e.g., *pinza* "pincer", *orzo* "barley", *colza* "rapeseed"). Participants were instructed to familiarise themselves with the sentence list and then to read it aloud in the most natural way. Each speaker recorded the sentence reading task in isolation, in order to avoid possible influences on other participants.

After the reading task, speakers of the same region and gender performed the second task, consisting of a map task dialogue. Logistic constraints prevented us from achieving a fully balanced design with equal numbers of male and female speakers from each geographical region. The map tasks elicited 38 target items, a subset of which also appeared in the reading task and were distributed across phonological contexts: 8 words were selected for the initial position, 18 for the intervocalic one, and 12 for the post-sonorant context. The maps presented both the image and an orthographic transcription in order to avoid the use of synonyms for the same target object. Each pair of speakers was presented with 4 maps, and speakers alternated in the role of "giver" of instruction and of "follower". In order to ensure that both speakers have access to the same items, target words have been balanced in the maps. However, since this task involves a certain degree of spontaneity, it was not possible to ensure that all the speakers would have produced the exact same number of items. Indeed, a certain degree of variability was expected, particularly for what it concerned repetitions of the same target words.

3.3. Data Collection and Annotation

The recordings were performed in a soundproof music room at the University College of Pavia Giasone del Maino, where all participants resided, using a Tascam recorder equipped with two dynamic microphones (frequency of sampling of the signal at 44.1 KHz and sampling rate of 16 bits). As stated before, both tasks were carried out by each participant on the same day, always following the order of reading tasks (first speaking 1, then speaking 2) and map tasks (speakers 1 and 2 together). The recordings lasted 18′37″ for the reading task (with a mean duration around 4′) and 4h4′8″ for the map-task dialogues (with a mean duration of 1 h for each couple). All data were transcribed and annotated using a semi-automatic procedure on the ELAN version 6.8 software using BAS—Web MAUS Basic and a forced alignment that creates a TextGrid executable in Praat. The corpus is composed of 408 tokens for the reading task (51 tokens × 8 speakers) and 1895 tokens for the map task, averaging 237 tokens per speaker. However, 110 tokens from the map-task dialogue were excluded due to voice overlap or external noise that disturbed the audio signal.

The TextGrid generated in ELAN was then imported into Praat for further processing. The file originally contained a single orthographic (ORT) tier, which was manually revised and corrected whenever audio–text alignment proved inaccurate. Because the ORT tier included the transcriptions of both speakers in the map-task dialogues, an additional

speaker tier (SP) was manually inserted to distinguish their productions. In accordance with the protocol described in Meluzzi (2020), we also added the tiers word, phone, and affricate. The Word tier was used to record the orthographic form of each token and to flag any realisations that differed from the intended target; such cases were marked as not allowed (e.g., zanzara_NA). The phone tier identified each dental affricate and labelled it according to three voicing categories based on acoustic criteria. Voicing classification was determined by measuring the proportion of the affricate's total duration displaying a visible voice bar (periodic low-frequency energy below 500 Hz) in the spectrogram. The stop and fricative portions of the affricate were distinguished by the presence of the release burst. Affricates were classified as:

- "AFFR+" (voiced): voice bar present for ≥75% of the total affricate duration;
- "AFFR−" (voiceless): voice bar present for <25% of the total affricate duration;
- "AFFR_MIX" (intermediate voicing): voice bar present for 25–75% of the total affricate duration.

Crucially, in intermediate realisations, voicing was consistently realised only during the stop portion, while the fricative portion (following the release burst) was invariably voiceless. As illustrated in Figure 1, the voicing bar (as well as the periodicity in the waveform) is clearly present in the occlusive portion of the affricate, but it progressively loses its periodicity towards the end of the occlusion, which is also characterised by multiple bursts; the following fricative section is clearly and completely voiceless. This acoustic pattern mirrors that documented in Meluzzi (2020), where the same annotation system was employed and the same distribution of voicing across the two portions of the affricate was observed. The voicing thresholds adopted here (75% and 25%) were also validated perceptually in Meluzzi (2020), where a preliminary perception study demonstrated that affricates with intermediate voicing degrees were categorised by listeners as neither fully voiced nor fully voiceless. All voicing measurements were conducted manually by the second author in Praat, with boundaries determined by visual inspection of both waveform and spectrogram.

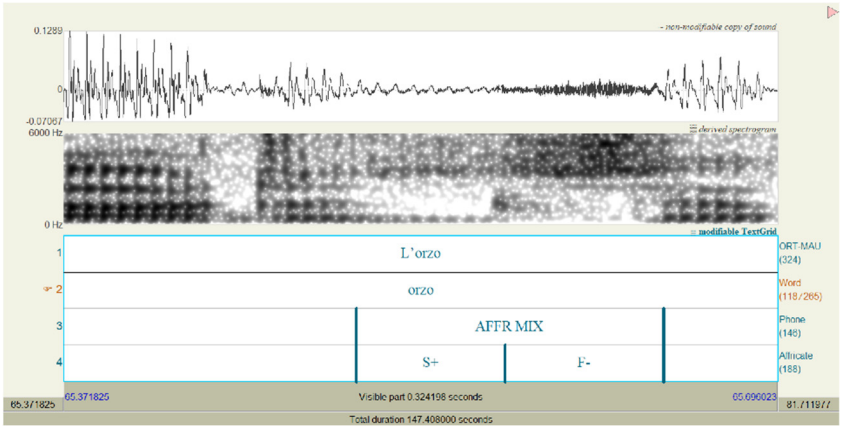

Figure 1. Instance of an alveolar affricate realised with an intermediate voicing degree, i.e., with voicing only in the occlusive portion.

In the intervocalic context, we also annotated the preceding vowel, with boundaries determined by the F2 and the waveform. In the initial context, the beginning of the affricate was identified in correspondence with the periodicity of the waveform in the case of sound realisation; however, for voiceless realisations, the boundary was set 50 ms before the burst, as in the CLIPS protocol (Crocco, 2001). The final vowel of the preceding word was also annotated for word-initial affricates; in this case, word-final vowels' label presents the symbol # (e.g., a#). It is worth noting that word-initial affricates were coded as such regardless

of whether they were preceded by a vowel-final word in connected speech (annotated with # for descriptive purposes). We did not separately code for *raddoppiamento fonosintattico* (RF), as our focus was on lexical frequency effects rather than on the variability introduced by specific syntactic contexts. While RF-triggered gemination could potentially affect duration measurements in word-initial position, previous studies on Italian dental affricates (Meluzzi, 2016, 2020) have shown that the voicing contrast is maintained independently of such contextual lengthening effects. Phonological context (word-initial vs. intervocalic, including lexical geminates) was included in our statistical models as a control variable.

The fourth and final tier affricate indicated the different phases of occlusion "S" and fricative moment "F", respectively, followed by the symbols "+" and "−" depending on the degree of voicing.

The acoustic analysis identified productions characterised by an initial voiced stop portion (S+) followed by a voiceless fricative portion (F−). These realisations, which Canepari (1997) termed intermediate, have been documented acoustically in several subsequent studies, including Meluzzi (2016) and Nese (2023). In the present annotation protocol, these tokens were labelled "MIX" in the Phone tier. Following Foulkes et al. (2011), instances of post-burst aperiodicity, characterised by a period of silence between the stop and fricative portions, were labelled "E" as a separate segment in the Phone tier. This aperiodic phase was observed in a small subset of affricates in the corpus; when present, the temporal sequence of affricate portions was S–E–F (see also Meluzzi, 2020). An additional phenomenon annotated in this tier concerns pre-aspiration, as defined by Stevens and Hajek (2007, 2010), which was coded as "hC" immediately following the stop label when present (e.g., "S– hC", indicating a voiceless affricate with pre-aspiration). Although the annotation protocol included categories for post-burst aperiodicity (E) and pre-aspiration (hC), these features were rare in the present dataset (fewer than 5% of tokens) and were not included in subsequent statistical analyses.

Durational values were extracted using a Praat script for the word, the dental affricate, and the distinct portions of the affricate. Affricate duration was included in the data matrix both as absolute duration (in milliseconds) and as normalised values. To account for individual differences in overall speech rate, affricate durations were z-score normalised speaker-wise following a procedure slightly different from the one described in Jacewicz et al. (2009), that was our main reference. In our case, for each speaker, z-scores were calculated as $z = (x - \mu\_speaker)/\sigma\_speaker$, in which x is the raw duration value, μ_speaker is the speaker's mean affricate duration across all tokens, and σ_speaker corresponds to the standard deviation. This speaker-specific normalisation recentres and rescales duration values based on each individual's baseline speaking characteristics, facilitating comparisons across speakers while preserving within-speaker variation.

Further linguistic information included in the matrix concerned lexical frequency and word etymology, since the latter is deeply linked to the expected voicing degree, as explained before. We firstly used the GRADIT (*Grande dizionario italiano dell'uso*) to obtain categorical information on both word frequency and etymology. Frequency was classified categorically regarding the frequency of use of the word, with the following labels: fundamental (FO it. *fondamentale*), high use (AU, it. *di alto uso*), high availability (AD, it. *di alta disponibilità*), common (CO, it. *comune*), specialised technical (TS, it. *tecnico specialistico*), and of regional use (RE, it. *uso regionale*). The fundamental, highly used, and highly available vocabularies together constitute the basic vocabulary of the Italian language. Since the target words also included personal names and city names, which were not provided in the GRADIT, they were indicated in the matrix as NP (it. *nome di persona*) and NC (it. *nome di città*), respectively. Regarding the etymology, the information in the GRADIT allowed a two-step classification: a first, more synthetic one indicated

the etymological language of the target word, while a more fine-grained classification also considered possible intermediate steps in the adoption of the loanword in the Italian lexicon. For instance, words that share the same etymological language may have arrived in Italian directly (e.g., Arabic > Italian) or through the mediation of an additional language (e.g., Arabic > Spanish > Italian). For some words, the etymology is unknown. For others, several hypotheses have been advanced, but none is predominant; thus, the word has been classified as having uncertain etymology.

Lexical frequency was quantified using the itTenTen20 corpus (Jakubíček et al., 2013), comprising approximately 3 billion tokens of Italian web text, accessed via Sketch Engine. The search was conducted using the Wordlist function, which allows for the extraction of lists of word forms or lemmas according to specific criteria. We employed the BASIC interface, selecting the following: the words category, which returns the actual orthographic forms attested in the corpus (rather than lemmas or other normalised linguistic units); and the containing filter option, in order to retrieve all forms containing the specified character string. The query yielded the absolute frequency of the target form in the corpus. The frequency value of each target word is also listed in the Supplementary Materials. To meet the requirements of a robust statistical analysis, we have transformed these raw values through the Zipf scale (van Heuven et al., 2014), calculated as $\log_{10}((\text{frequency}/3{,}000{,}000{,}000) \times 1{,}000{,}000 + 1) + 3$, to meet distributional assumptions for linear modelling and appropriately handle zero-frequency items (n = 201, 8.5% of the dataset). The Zipf scale typically ranges from 3 (very low frequency) to 7 (very high frequency), thus meeting distributional assumptions for linear modelling while appropriately handling zero-frequency items. The Zipf scale is widely adopted in psycholinguistic research as it provides a more interpretable metric than raw logarithmic transformations and accounts for the highly skewed distribution characteristic of word frequency data, as in our dataset.

4. Analysis

4.1. Voicing Degree

As specified in Section 3.2, we have spectrographically identified three different degrees of affricate voicing, namely voiced, voiceless, and intermediate (see also Meluzzi, 2020). In the following analysis, we have firstly checked if the distribution of the voicing degrees correlates with the possible linguistic variables expected from literature (see Meluzzi, 2016, 2020).

Initial investigation of the whole corpus confirms that affricate variants are distributed according to the phonological context (Table 1). Voiced affricates occur predominantly in word-initial position (96.1%), while intervocalic geminates and post-lateral affricates are most often realised as voiceless (77.5% and 84.5%, respectively). By contrast, the post-nasal context exhibits the highest proportion of voiced realisations (56.1%), followed by post-rhotic (39.5%) and intervocalic singletons (35.8%). Intermediate affricates are relatively rare across all contexts and appear primarily in word-initial and intervocalic geminate positions (respectively, 11 and 10 cases out of 37 total intermediate affricates).

As expected from previous works (cf. Section 2.2), speakers' geographical origin affects affricate voicing across phonological contexts (Table 2). Northern and southern speakers display broadly similar patterns in voiceless affricate production, with comparable distributions across contexts. Word-initial affricates are predominantly realised as voiced by both groups; however, southern speakers also produce approximately 25% of voiceless variants in this context.

Table 1. Voicing distribution across phonological contexts in the corpus (n = 1735).

	Initial	Intervocalic Geminate	Intervocalic Singleton	Post-Lateral	Post-Rhotic	Post-Nasal
Voiceless	7 (1.5%)	457 (77.5%)	117 (62.7%)	60 (84.5%)	105 (59.3%)	100 (40.7%)
Voiced	446 (96.1%)	123 (20.8%)	67 (35.8%)	8 (11.3%)	70 (39.5%)	138 (56.1%)
Intermediate	11 (2.4%)	10 (1.7%)	3 (1.6%)	3 (4.2%)	2 (1.1%)	8 (3.3%)
Total	464 (100%)	590 (100%)	187 (100%)	71 (100%)	177 (100%)	246 (100%)

Table 2. Affricate realisations divided between northern and southern speakers across phonological contexts.

		Initial	Intervocalic Geminate	Intervocalic Singleton	Post-Lateral	Post-Rhotic	Post-Nasal
Northern Speakers	**Voiceless**	12.2%	80.9%	44.2%	96.6%	49.3%	31.9%
	Voiced	85.1%	17.4%	55.2%	1.7%	50.0%	56.2%
	Intermediate	2.7%	1.7%	0.6%	1.7%	0.7%	11.9%
Southern Speakers	**Voiceless**	25.7%	84.2%	46.3%	75.7%	46.3%	29.9%
	Voiced	73.7%	15.0%	51.6%	18.9%	52.5%	65.0%
	Intermediate	0.6%	0.8%	2.1%	5.4%	1.3%	5.1%

Post-sonorant contexts exhibit the greatest geographical variability in our dataset, with important differences between post-lateral and post-nasal contexts. In the post-lateral position, northern speakers produce affricates almost exclusively as voiceless (96.6%), while southern speakers show a substantial proportion of voiced realisations (18.9%). In the post-nasal position, both northern and southern speakers favour the voiced variant, though voiceless realisations also occur with similar frequencies (31.9% for northern speakers, 29.9% for southern speakers). This suggests that voicing preferences in this context is more closely linked to lexical properties than to geographical origin.

The post-nasal context also exhibits the highest rate of intermediate realisations in our dataset. Moreover, within this context, northern speakers produce intermediate affricates more frequently than southern speakers (11.9% vs. 5.1%).

Analysis of voicing distribution as linked to the lexicon was performed by using GRADIT classification and etymology, as previously explained in Section 3.2. Results are shown in Tables 3 and 4.

Table 3. Affricate voicing and lexical categorical frequency (GRADIT category, with English translation).

	GRADIT Classification							
	***AD* (*High Availability*)**	***AU* (*High Use*)**	***CO* (*Common Use*)**	***FO* (*Fundamental*)**	***NC* (*City Name*)**	***NO* (*Personal Name*)**	***RE* (*Regional Use*)**	***TS* (*Specialised Technical*)**
Voiceless	28.4%	57.9%	58.2%	61.8%	35.0%	32.5%	54.2%	39.7%
Voiced	69.7%	40.5%	38.0%	38.2%	63.1%	65.0%	39.6%	60.3%
Intermediate	1.8%	1.7%	3.8%	0.0%	1.9%	2.5%	6.3%	0.0%

Table 4. Affricate voicing and word etymology.

	Etymology								
	Uncertain	***Arabic***	***French***	***Germanic***	***Japanese***	***Greek***	***Latin***	***Dutch***	***Polish***
Voiceless	57.8%	29.7%	88.1%	21.3%	7.3%	37.7%	59.6%	100.0%	0.0%
Voiced	41.0%	67.0%	4.8%	77.5%	92.7%	62.3%	37.6%	0.0%	96.6%
Intermediate	1.1%	3.3%	7.1%	1.3%	0.0%	0.0%	2.7%	0.0%	3.4%

In line with previous accounts of the diachronic development of Italian affricates (Celata, 2004; Meluzzi, 2021), voiceless realisations are especially common in words of

Romance origin in our data, such as those inherited from Latin or borrowed from French, and also in the (albeit very few) borrowings from Dutch. Within our dataset, core lexical items in Italian likewise tend to favour the voiceless variant. By contrast, voiced affricates are more frequently found in words of limited lexical availability, including proper nouns denoting places and individuals, as well as in loanwords from non-Romance languages. Our preliminary descriptive analysis confirmed that voicing distribution in our corpus aligns with these diachronic patterns. Of the 2193 dental affricate tokens analysed, 50.7% (n = 1111) were realised as voiceless, 46.6% (n = 1022) as voiced, and 2.7% (n = 60) as intermediate. Words of Romance origin (particularly from Latin etyma) showed a strong preference for voiceless variants, whilst words from Arabic and other non-Romance sources (with the exception of Dutch) favoured voiced realisations. Crucially, intermediate realisations (which represent neither a voiceless nor a fully voiced articulation) were not predicted by any specific etymological origin, suggesting they may reflect autonomous sociophonetic developments or context-dependent gradient phonetic realisations (see also Meluzzi, 2020).

The overall voicing distribution observed in our data aligns with previous findings in the literature. We now turn to the specific role of lexical frequency in conditioning dental affricate realisation. Given that both voiceless and voiced outcomes represent legitimate historical developments based on etymology, and that intermediate realisations constitute a qualitatively distinct category, we employ a multinomial logistic mixed-effects framework. Voiceless realisations were selected as the reference category on both empirical grounds (being the most frequent outcome, 50.7%) and theoretical grounds (allowing direct comparison of voiced and intermediate variants against the historically unmarked voiceless outcome).

To examine the effects of lexical frequency on affricate voicing, we fitted a multinomial mixed-effects regression model using the mblogit function from the mclogit package (Elff, 2022) in R 4.5.2. This implementation was preferred over standard multinomial logistic regression (e.g., nnet::multinom) as it explicitly supports crossed random intercepts within a multinomial framework, which is essential given the nested structure of our data. Random intercepts were included for speaker (n = 8) and item (n = 60 unique lexical types, as a subset of items appeared in both tasks) to account for the non-independence of observations arising from multiple tokens produced by the same speakers and representing repeated measurements of the same lexical items. The effects of lexical frequency on affricate duration are examined separately in Section 4.2 using a linear mixed-effects model with an analogous random effects structure.

Fixed effects included lexical frequency (Zipf-transformed, mean-centred, as explained in Section 3), geographical origin (north vs. south), phonological context (with six levels: intervocalic geminate, intervocalic singleton, initial, post-nasal, post-rhotic, and post-lateral), and speech style (dialogue vs. reading). Voicing was treated as a three-level non-ordinal outcome (i.e., voiceless, voiced, and intermediate). Voiceless was set as the reference category on both empirical grounds (being the most frequent outcome, 50.7%) and theoretical grounds, thus allowing direct comparison of voiced and intermediate variants against the historically unmarked voiceless outcome.

We first fitted a baseline model including main effects of all predictors:

$$Voicing \sim ItemFreq\_Zipf + GeoOrigin + PHON + Style + (1 \mid Speaker) + (1 \mid Item)$$

We then tested whether the effect of lexical frequency varied by geographical origin by including their interaction to determine whether frequency effects on voicing varied across northern and southern speakers:

$$Voicing \sim ItemFreq\_Zipf \times GeoOrigin + PHON + Style + (1 \mid Speaker) + (1 \mid Item)$$

Model comparison via likelihood ratio test indicated that the interaction did not significantly improve model fit ($\chi^2(2) = 1.88$, $p = 0.391$), and the baseline model was therefore retained for subsequent interpretation. The baseline model achieved a classification accuracy of 91.7% on the training data. Inspection of the confusion matrix revealed that the model successfully distinguished voiced (91.9% correct) and voiceless realisations (96.4% correct) but failed to identify intermediate realisations as a distinct category, consistently classifying them as either voiced (n = 27) or voiceless (n = 33). This is consistent with the small proportion of intermediate tokens in the dataset (2.7%, n = 60) and the non-significant coefficient for this category, suggesting that intermediate realisations may not constitute a sufficiently distinct acoustic category to be reliably identified with the current sample size (see also the theoretical discussion on this category in Meluzzi, 2020).

The results of the baseline model are summarised in Table 5. Lexical frequency does not reach statistical significance as a predictor of voicing category (voiced vs. voiceless: β = −0.772, SE = 0.592, OR = 0.462, 95% CI [0.145, 1.475], $p = 0.192$; intermediate vs. voiceless: β = −0.727, SE = 0.406, OR = 0.483, 95% CI [0.218, 1.072], $p = 0.073$), suggesting that once between-speaker and between-item variability are appropriately accounted for, lexical frequency does not independently determine voicing category.

Table 5. Results of the multinomial mixed-effects model (mblogit; Elff, 2022).

Predictor	Comparison	β	SE	OR	95% CI	*p*
Lexical Frequency (Zipf)	Voiced vs. Voiceless	−0.772	0.592	0.462	[0.145, 1.475]	0.192
	Intermediate vs. Voiceless	−0.727	0.406	0.483	[0.218, 1.072]	0.073
Geographical Origin (South)	Voiced vs. Voiceless	+0.569	0.231	1.766	[1.124, 2.775]	0.014 *
	Intermediate vs. Voiceless	+0.426	0.388	1.531	[0.716, 3.278]	0.272
Phonological Context (ref. = Geminate)						
Initial	Voiced vs. Voiceless	+2.099	0.958	8.154	[1.248, 53.295]	0.028 *
	Intermediate vs. Voiceless	+1.523	0.725	4.585	[1.108, 18.973]	0.036 *
Singleton	Voiced vs. Voiceless	−0.917	1.476	0.400	[0.022, 7.219]	0.535
	Intermediate vs. Voiceless	−0.940	1.020	0.391	[0.053, 2.888]	0.357
Post-Nasal	Voiced vs. Voiceless	+0.900	1.318	2.458	[0.186, 32.578]	0.495
	Intermediate vs. Voiceless	+0.198	0.874	1.219	[0.220, 6.758]	0.821
Post-Rhotic	Voiced vs. Voiceless	+0.022	1.882	1.022	[0.026, 40.919]	0.991
	Intermediate vs. Voiceless	−0.431	1.235	0.650	[0.058, 7.315]	0.727
Post-Lateral	Voiced vs. Voiceless	−0.745	1.807	0.475	[0.014, 16.385]	0.680
	Intermediate vs. Voiceless	−0.923	1.206	0.397	[0.037, 4.227]	0.444
Speech Style (Reading)	Voiced vs. Voiceless	−0.462	0.292	0.630	[0.355, 1.118]	0.114
	Intermediate vs. Voiceless	+0.573	0.343	1.774	[0.905, 3.474]	0.095

Note. Reference categories: voiceless (outcome), geminate (phonological context), north (geographical origin), and dialogue (speech style). Random intercepts included for speaker (n = 8) and word (n = 60). * $p < 0.05$.

Phonological context emerges as the strongest predictor of voicing patterns. Word-initial positions show the most pronounced effect, strongly favouring both voiced (β = +2.099, SE = 0.958, OR = 8.154, 95% CI [1.248, 53.295], $p = 0.028$) and intermediate realisations (β = +1.523, SE = 0.725, OR = 4.585, 95% CI [1.108, 18.973], $p = 0.036$) relative to the geminate reference context. Whilst the wide confidence intervals for the initial context reflect the inherent variability of word-initial affricates in connected speech, where factors such as prosodic boundary strength, speech rate, and preceding context may modulate realisation, the effect is robust and consistent with the literature documenting word-initial position as the primary locus of voicing in Italian dental affricates (cf. Meluzzi, 2016, 2020). No other phonological context reaches significance, though the particularly wide confidence intervals for post-nasal, post-rhotic, and post-lateral contexts reflect the

comparatively smaller number of tokens in these categories and suggest that results should be interpreted with caution.

Geographical origin shows a significant effect on voiced versus voiceless realisations (β = +0.569, SE = 0.231, OR = 1.766, 95% CI [1.124, 2.775], p = 0.014), indicating that southern speakers produced voiced affricates more frequently than northern speakers after controlling for phonological context, lexical frequency, and speech style. Geographical origin has no significant effect on intermediate realisations (β = +0.426, SE = 0.388, p = 0.272). Finally, speech style did not reach significance for either comparison (voiced vs. voiceless: β = −0.462, p = 0.114; intermediate vs. voiceless: β = +0.573, p = 0.095), though a trend towards more intermediate realisations in reading style warrants further investigation in future work.

Figure 2 displays the predicted probabilities of voiceless, voiced, and intermediate realisations as a function of lexical frequency (Zipf scale), separately for northern and southern speakers; predictions are estimated from the baseline model, holding phonological context at intervocalic geminate and speech style at dialogue. In both groups, the probability of voiceless realisations increases monotonically with lexical frequency, whilst the probability of voiced realisations decreases correspondingly. The geographical origin effect is visible in the different starting points of the curves at low frequency values (Zipf ≈ 3): southern speakers show a substantially higher predicted probability of voiced realisations (≈45%) compared to northern speakers (≈30%). This is consistent with the significant geographical origin effect reported in Table 5 (β = +0.569, p = 0.014). Moreover, intermediate realisations remain consistently low across the entire frequency range in both groups, converging towards zero at high frequency values.

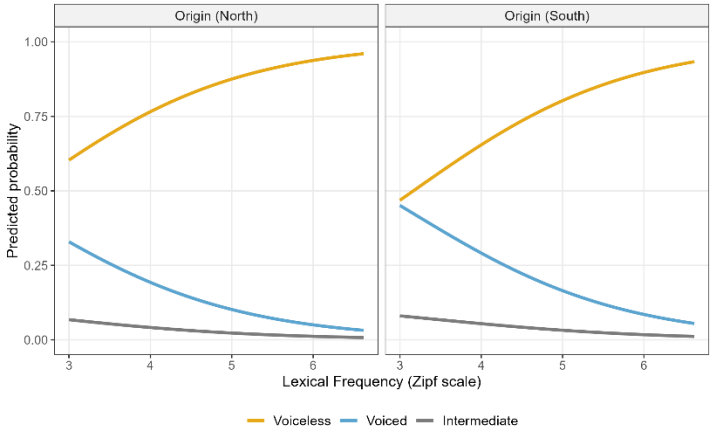

Figure 2. Predicted probability of affricate voicing realisations as a function of lexical frequency (Zipf scale) for northern and southern speakers.

Figure 3 illustrates the predicted voicing probabilities across the six phonological contexts; in this case, predictions are estimated from the baseline model, holding geographical origin at north, and speech style at dialogue, and intervocalic geminate serves as the reference context. Whilst the general pattern of increasing voiceless probability with lexical frequency is consistent across all contexts, the starting point of the curves varies substantially, reflecting the strong main effect of phonological context reported in Table 5. The most striking contrast emerges between word-initial position and all other contexts: in the initial position, voiced realisations are strongly favoured at low frequency values (predicted probability ≈ 75%), with voiceless realisations becoming dominant only at higher frequency values, where the two curves intersect. This pattern is unique to word-initial position and is consistent with the significant positive coefficient reported for this context (β = +2.099, p = 0.028). In the intervocalic geminate context, voiceless realisations are

strongly dominant across the entire frequency range, with voiced realisations remaining below 35% even at the lowest frequency values. The Intervocalic Singleton and Post-Lateral contexts pattern similarly to Intervocalic Geminate, showing consistently high probabilities of voiceless realisations throughout. Post-Nasal context shows an intermediate pattern, with voiced realisations slightly favoured at low frequency values before giving way to voiceless realisations at higher frequencies, though this effect did not reach significance in the model (β = +0.900, p = 0.495). Intermediate realisations remain consistently low across all contexts and frequency values, further corroborating the confusion matrix results discussed above.

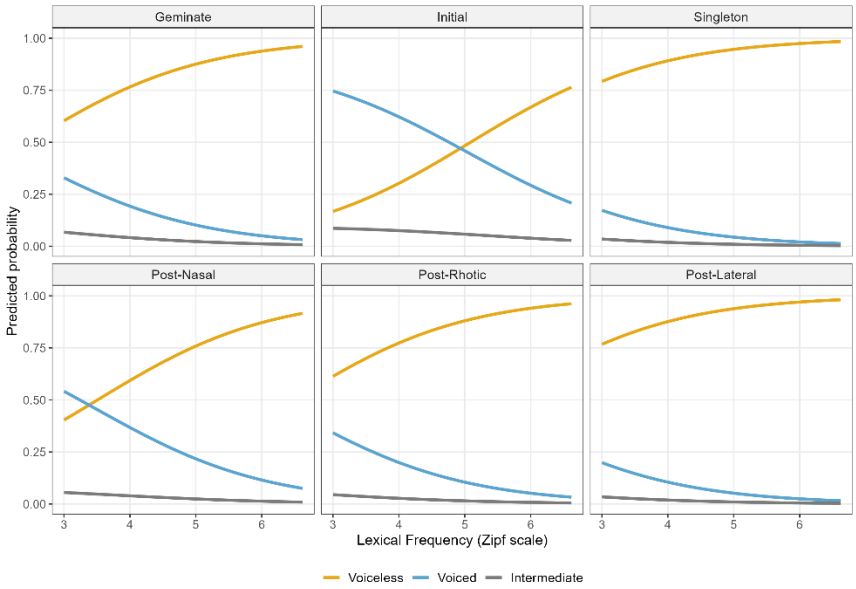

Figure 3. Predicted probability of affricate voicing realisations as a function of lexical frequency (Zipf scale) across phonological contexts.

4.2. Duration

Prior to model fitting, we examined the bivariate relationships between the key continuous variables (namely, affricate duration and word duration). This preliminary step was motivated by two considerations. First, the moderate correlation between affricate and word duration (r = 0.455, $p < 0.001$) raised the possibility that any frequency effect on affricate duration might be partially mediated by frequency-driven reduction at the word level rather than reflecting a segment-specific process (as well-documented since Gahl, 2008, but with opposing results in the literature). Second, given that the voicing category had been treated as the dependent variable in Section 4.1, its potential influence on affricate duration warranted explicit examination before including it as a covariate. Indeed, voiced affricates are known to differ systematically from voiceless ones in their temporal properties (cf. Meluzzi, 2016, 2020). Including both normalized word duration (Word_Dur_z) and voicing category as covariates in the duration model allowed us to isolate the independent contribution of lexical frequency to affricate duration whilst appropriately controlling for these sources of variation.

Lexical frequency showed a weak but significant positive correlation with affricate duration (r = 0.058, p = 0.007) and a moderate negative correlation with word duration (r = −0.331, $p < 0.001$), the latter consistent with well-documented frequency-based reduction effects at the word level, as in the aforementioned Gahl's (2008) work.

4.2.1. Model Specification and Comparison

To examine the effects of lexical frequency on affricate duration, we fitted a series of linear mixed-effects models using the lmer function from the lme4 package (Bates et al., 2015) in R, with *p*-values obtained via Satterthwaite's method as implemented in lmerTest (Kuznetsova et al., 2017). The dependent variable was speaker-wise z-scored affricate duration (Affricate_Dur_z), following the normalisation procedure previously described in Section 3. All models included random intercepts for speaker (n = 8) and item (n = 60 unique lexical types, as a subset of items appeared in both tasks) to account for non-independence of observations. Fixed effects included lexical frequency (Zipf-transformed, mean-centred), geographical origin (north vs. south), phonological context (six levels: intervocalic geminate, intervocalic singleton, initial, post-nasal, post-rhotic, and post-lateral), and speech style (dialogue vs. reading). From Step 2 onwards, word duration (z-scored, speaker-wise) and voicing category (voiceless, voiced, and intermediate) were also added to the model. All models were fitted using maximum likelihood (REML = FALSE) for model comparison purposes. Then, the final model was refitted using restricted maximum likelihood (REML = TRUE) for parameter estimation.

Model selection proceeded in four steps, each evaluated via a likelihood ratio test (LRT) against the preceding model:

Step 1. Baseline model with main effects only:

$$Duration \sim ItemFreq\_Zipf + GeoOrigin + PHON + Style + (1 \mid Speaker) + (1 \mid Word)$$

Step 2. Addition of word duration as covariate:

$$Duration \sim ItemFreq\_Zipf + GeoOrigin + PHON + Style + Word\_Dur\_z + (1 \mid Speaker) + (1 \mid Word)$$

Step 3. Addition of voicing category as covariate:

$$Duration \sim ItemFreq\_Zipf + GeoOrigin + PHON + Style + Word\_Dur\_z + Affricate\_clean + (1 \mid Speaker) + (1 \mid Word)$$

Step 4. Two interaction terms were tested against Step 3, separately:

$$Duration \sim ItemFreq\_Zipf \times GeoOrigin + PHON + Style + Word\_Dur\_z + Affricate\_clean + (1 \mid Speaker) + (1 \mid Word)$$

$$Duration \sim ItemFreq\_Zipf \times PHON + GeoOrigin + Style + Word\_Dur\_z + Affricate\_clean + (1 \mid Speaker) + (1 \mid Word)$$

The results of the model comparison are summarised in Table 6. For each step, the number of parameters (k), AIC, log-likelihood (−2LL), likelihood ratio statistic (χ^2), degrees of freedom (df), and *p*-value are reported.

Table 6. Summary of model comparison for affricate duration.

Step	Model Specification	k	AIC	−2LL	χ^2	df	*p*
1	Main effects only	12	4476.8	4452.8	—	—	—
2	+Word_Dur_z	13	3304.2	3278.2	1174.6	1	<0.001 ***
3	+Voicing	15	3199.7	3169.7	108.5	2	<0.001 ***
4a	+Freq × GeoOrigin	16	3201.5	3169.5	0.2	1	0.679
4b	+Freq × PHON	20	3203.7	3163.7	6.0	5	0.309

Note. k = number of parameters; AIC = Akaike Information Criterion; −2LL = −2 log-likelihood; χ^2 = likelihood ratio test statistic; df = degrees of freedom. Each model is tested against the preceding model in the sequence. Model 3 (main effects + Word_Dur_z + Voicing) was selected as the final model. *** $p < 0.001$.

The inclusion of word duration significantly improved model fit ($\chi^2(1) = 1174.6$, $p < 0.001$), as did the addition of voicing category ($\chi^2(2) = 108.45$, $p < 0.001$). Neither the interaction between lexical frequency and geographical origin ($\chi^2(1) = 0.171$, $p = 0.679$) nor that between lexical frequency and phonological context ($\chi^2(5) = 5.967$, $p = 0.309$) significantly improved model fit, and both interaction terms were therefore excluded from the final model. The final model thus included main effects of all predictors together with word duration and voicing category as covariates.

4.2.2. Results

The results of the final model are reported in Table 7. The model was fitted with REML = TRUE and included random intercepts for word ($\sigma^2 = 0.246$) and speaker ($\sigma^2 = 0.003$), with residual variance $\sigma^2 = 0.227$. The comparatively low variance attributed to the speaker is expected, given that affricate durations were z-score normalised speaker-wise prior to analysis.

Table 7. Results of the linear mixed-effects model for affricate duration (z-score).

Predictor	β	SE	t	*p*
Lexical Frequency (Zipf)	+0.153	0.076	2.018	0.050 *
Geographical Origin (South)	−0.091	0.045	−2.004	0.088
Word Duration (z-score)	+0.588	0.015	39.925	<0.001 ***
Voicing				
Voiced vs. Voiceless	−0.484	0.047	−10.334	<0.001 ***
Intermediate vs. Voiceless	−0.099	0.069	−1.427	0.154
Phonological Context				
Initial	−0.184	0.118	−1.555	0.121
Singleton	+0.197	0.181	1.088	0.282
Post-Nasal	−1.003	0.130	−7.715	<0.001 ***
Post-Rhotic	−0.527	0.243	−2.166	0.036 *
Post-Lateral	−0.187	0.159	−1.176	0.241
Speech Style (Reading)	+0.135	0.031	4.383	<0.001 ***
Random Effects	**Variance**	**SD**		
Word (Intercept)	0.246	0.496		
Speaker (Intercept)	0.003	0.056		
Residual	0.227	0.477		

Note. Dependent variable: z-score normalised affricate duration (speaker-wise). Reference categories: voiceless (voicing), geminate (phonological context), north (geographical origin), dialogue (speech style). Random intercepts included for word (n = 60) and speaker (n = 8). n = 2192 observations. * $p < 0.05$; *** $p < 0.001$.

Lexical frequency emerged as a significant positive predictor of affricate duration (β = +0.153, SE = 0.076, t = 2.018, $p = 0.050$). Therefore, once word duration and voicing category are controlled for, higher-frequency items are associated with longer affricate durations. The implications of this finding are discussed in Section 5.

Phonological context was a strong predictor of duration. Post-nasal context shows the largest effect, with affricates in this position being substantially shorter than intervocalic geminates (β = −1.003, SE = 0.130, t = −7.715, $p < 0.001$). Post-rhotic context also shows significantly shorter durations relative to the intervocalic geminate used as reference (β = −0.527, SE = 0.243, t = −2.166, $p = 0.036$). Statistically significant results are plotted and described in Figure 4 below. It is worth noting that no other phonological context

reached significance: this means that affricates in initial (β = −0.184, SE = 0.118, p = 0.121), intervocalic singleton (β = +0.197, SE = 0.181, p = 0.282), and post-lateral (β = −0.187, SE = 0.159, p = 0.241) contexts do not differ significantly from intervocalic geminate after controlling for word duration and voicing.

As expected, the voicing category was a significant predictor of duration. Voiced affricates are substantially shorter than voiceless ones (β = −0.484, SE = 0.047, t = −10.334, $p < 0.001$): this result is consistent with well-established findings on the durational correlates of laryngeal contrasts in Italian affricates (Meluzzi, 2016, 2020). Furthermore, intermediate realisations did not differ significantly from voiceless ones (β = −0.099, SE = 0.069, p = 0.154). Speech style showed a significant effect, with affricates in reading style being longer than in dialogue (β = +0.135, SE = 0.031, t = 4.383, $p < 0.001$): this is also consistent with a major control on form typically associated with formal registers. Interestingly, geographical origin did not reach significance (β = −0.091, SE = 0.045, p = 0.088), though a trend towards shorter durations in southern speakers was observed.

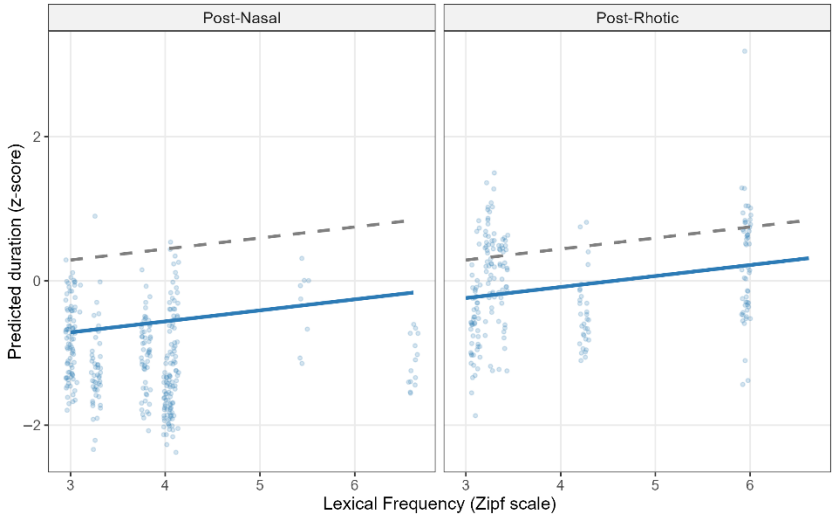

Figure 4. Predicted affricate duration (z-score) as a function of lexical frequency (Zipf scale) for post-nasal and post-rhotic phonological contexts.

Figure 4 displays the predicted affricate duration (z-score) as a function of lexical frequency for the two phonological contexts that reached significance in the model, Post-Nasal and Post-Rhotic, plotted against the Geminate reference context (dashed line). Raw data points are shown in the background to convey the empirical distribution of observations. It should be noted that tokens in both contexts are predominantly concentrated in the lower frequency range (Zipf ≈ 3–4), with comparatively few observations at higher frequency values; predicted values beyond this range should therefore be interpreted with caution. Notwithstanding this limitation, both contexts show consistently shorter predicted durations relative to the Geminate reference across the entire frequency range, with Post-Nasal showing a substantially larger offset (β = −1.003) than Post-Rhotic (β = −0.527).

5. Discussion

Before discussing the substantive findings, it is necessary to foreground several methodological constraints that delimit the interpretive reach of this study. First and most notably, the sample comprised only eight speakers, evenly stratified by geographical origin and sex. Whilst this sample size is comparable to other previous phonetic studies examining fine-grained segmental variation in spontaneous speech (for instance, Meluzzi, 2016, 2020), it obviously limits our ability to generalise across the broader Italian-speaking population and precludes detailed investigation of individual differences. Second, the

distribution of lexical frequency across phonological contexts was uneven, with tokens in post-nasal and post-rhotic positions predominantly concentrated in the lower frequency range (Zipf ≈ 3–4) and comparatively few observations at higher frequency values. However, this distribution reflects the structural properties of the Italian lexicon itself: indeed, dental affricates are a marked consonant class (in Italian as well as at the interlinguistic level, cf. Ladefoged & Maddieson, 1996) with limited lexical distribution, and their occurrence in post-consonantal contexts is further constrained by phonotactic restrictions. As such, this limitation is not readily avoidable without resorting to nonce words, which would introduce their own confounds. Nevertheless, predicted values beyond the empirically attested frequency range should be interpreted with appropriate caution. Third, intermediate voicing realisations represented only 2.7% of the dataset and were not reliably classified by the model. However, this finding is consistent with aforementioned research on Italian dental affricates (e.g., Meluzzi, 2020; Nese & Meluzzi, 2017), which has documented similarly low proportions of intermediate tokens and suggested that they may arise from accommodation processes in dialogue rather than constituting a stable phonetic category. Taken together, while these limitations constrain the generalisability of certain findings, the robustness of the main effects reported here, in particular those involving phonological context, voicing category, and word duration, provides confidence in the validity of the core results.

In this paper, we have shown how, as expected from previous research on Italian dental affricates (Meluzzi, 2016, 2020), phonological context emerged as the strongest predictor of both voicing and duration patterns. Word-initial position strongly favoured voiced realisations ($\beta = +2.099$, $p = 0.029$), consistent with the well-documented tendency for Italian dental affricates to undergo voicing in this context across regional varieties. Post-nasal and post-rhotic contexts were characterised by substantially shorter affricate durations ($\beta = -1.003$ and $\beta = -0.527$, respectively), reflecting assimilatory or coarticulatory influences from the preceding sonorant. These findings align with theoretical expectations derived from both articulatory phonetics and phonological markedness hierarchies, reinforcing the robustness of context-specific phonetic implementations across speakers.

Geographical origin showed a significant effect on voicing category, with southern speakers producing voiced affricates more frequently than northern speakers ($\beta = +0.569$, $p = 0.014$). This pattern corroborates earlier observations (Meluzzi, 2020) documenting regional variation in the distribution of dental affricate voicing, particularly in post-sonorant contexts where southern varieties favour voiced variants. Interestingly, however, geographical origin did not reach significance as a predictor of affricate duration ($\beta = -0.091$, $p = 0.088$), suggesting that while regional varieties differ in their preference for voicing categories, they do not systematically differ in the temporal implementation of those categories once voicing itself is controlled for. This dissociation between categorical and gradient levels of representation is theoretically significant and foreshadows the more general pattern discussed in Section 5.1.

Finally, the voicing category emerged as a robust predictor of duration, with voiced affricates being substantially shorter than voiceless ones ($\beta = -0.484$, $p < 0.001$). This finding is consistent with well-established cross-linguistic patterns in the durational correlates of laryngeal contrasts and with previous work on Italian affricates specifically (Meluzzi, 2016, 2020), where voicing has been shown to exert systematic effects on segmental duration independently of other phonological and lexical factors.

5.1. Lexical Frequency Effects Between Categorical and Gradient Levels

The most theoretically consequential findings concern the role of lexical frequency in shaping affricate variation. Crucially, lexical frequency did not emerge as a significant

predictor of voicing category once between-speaker and between-item variability were appropriately accounted for through random effects (voiced vs. voiceless: $\beta = -0.772$, $p = 0.192$; intermediate vs. voiceless: $\beta = -0.727$, $p = 0.073$). This null result is theoretically significant in two respects. First, it contrasts with cross-linguistic tendencies documented for affricate voicing systems, where high-frequency items have been shown to favour articulatorily simpler variants, as recently shown by Dediu et al. (2024) for dental fricatives. Historical developments in Romance languages illustrate this pattern: Latin dental affricates /ts dz/ underwent widespread desonorisation and subsequent spirantisation in wide varieties, with voiced affricates either merging with voiceless counterparts or reducing to fricatives /s z/ (e.g., French and Spanish; Żygis et al., 2012, but cf. Meluzzi, 2021 for a review). The fact that lexical frequency does not systematically favour voiceless realisations in our dataset suggests that the maintenance of the voicing contrast in Italian dental affricates is not driven by frequency-conditioned articulatory simplification but rather is stabilised by other factors. Indeed, recent cross-linguistic research has documented frequency effects on consonant realisation in numerous languages, including English (Raymond et al., 2006), Dutch (Pluymaekers et al., 2005), Spanish (Torreira & Ernestus, 2011), Japanese (Hashimoto, 2021, 2023), and Mandarin (Tang & Shaw, 2021). Nevertheless, these effects typically manifest as gradient phonetic reduction rather than categorical shifts in voicing contrasts. The absence of a frequency effect on voicing category in Italian dental affricates thus aligns with broader cross-linguistic patterns, suggesting that lexical frequency primarily modulates fine-grained phonetic implementation rather than determining phonological categorisation.

Second, the null result corroborates previous findings by the present authors (Meluzzi, 2020), which demonstrated that the distribution of voiced and voiceless dental affricates in Italian is strongly modulated by sociolinguistic variables (thus including regional origin, speech style, and interactional context) rather than being determined solely by lexical properties. The substantial between-speaker variability observed in the present dataset (as reflected in the random intercepts for speaker: $\sigma^2 = 0.002$) supports this interpretation: the voicing of Italian dental affricates appears to constitute a sociolinguistic variable whose realisation is conditioned by speakers' regional and stylistic repertoires, with lexical frequency playing at most a marginal role once this sociolinguistic variation is appropriately modelled. This pattern stands in contrast to segmental variables such as vowel reduction or consonant lenition, where lexical frequency effects have been shown to operate more systematically across speakers and contexts (Bybee, 2001; Gahl, 2008).

Indeed, in our model, when between-speaker and between-item variability are properly accounted for, the systematic influence of lexical frequency on categorical voicing distinctions disappears. This finding has important theoretical implications: it suggests that lexical frequency does not directly determine phonemic categorisation but rather modulates phonetic implementation at a gradient level, as discussed below.

In contrast to the null result for voicing, in fact, lexical frequency emerged as a significant predictor of affricate duration, albeit with a counterintuitive positive coefficient ($\beta = +0.153$, $p = 0.050$): higher-frequency items were associated with longer rather than shorter affricate durations. This pattern initially appears to contradict well-established findings in the frequency-reduction literature (starting from Gahl, 2008), where high-frequency words are typically characterised by shorter segmental durations. However, the interpretation becomes clear when we consider that word duration was included as a covariate in the model. Lexical frequency showed a significant negative correlation with word duration ($r = -0.331$, $p < 0.001$), consistent with frequency-driven reduction at the word level. Once this word-level reduction is statistically controlled for, the affricate segment itself shows a positive relationship with frequency. This suggests a compensatory mechanism: while high-frequency words undergo overall temporal compression, the

affricate segment within these words is relatively preserved or even hyperarticulated, possibly to maintain segmental distinctiveness in the face of word-level reduction. This dissociation between word-level and segment-level durational effects aligns with well-known findings in the hyper- and hypoarticulation literature (Lindblom, 1990), where speakers balance articulatory economy with the need to preserve phonetic contrast.

Furthermore, the effect of lexical frequency on duration was uniform across phonological contexts. Although the interaction between frequency and phonological context (Freq × PHON) was tested, it did not significantly improve model fit ($\chi^2(5) = 5.967$, $p = 0.309$), indicating that the positive relationship between frequency and affricate duration operates consistently regardless of whether the affricate occurs in intervocalic geminate, intervocalic singleton, initial, or post-consonantal position. This uniformity contrasts with the strong main effect of phonological context itself: affricates in post-nasal and post-rhotic positions were substantially shorter than those in intervocalic geminate contexts, but the influence of lexical frequency on duration did not vary as a function of these contextual differences.

Taken together, these findings reveal a fundamental dissociation: lexical frequency does not determine categorical phonemic choices (voicing) but modulates gradient phonetic implementation (duration). This pattern is inconsistent with classical phonological models that assume a strict separation between lexical access and phonological rule application, where frequency effects would be expected to operate uniformly across both categorical and gradient levels of representation. Conversely, the present results support usage-based and exemplar-theoretic frameworks (Bybee, 2001, 2007; Pierrehumbert, 2002, 2016), in which phonetic variants are stored as probabilistic distributions shaped by cumulative exposure, and where lexical frequency influences the fine-grained phonetic detail of productions without necessarily determining higher-level categorical distinctions. The fact that frequency effects emerge for duration but not for voicing suggests that frequency-conditioned phonetic variation operates primarily at the sub-phonemic, implementation level, where speakers balance articulatory routinisation with the need to preserve perceptually relevant contrasts.

5.2. Theoretical Implications

The dissociation between categorical and gradient frequency effects documented in this work has broader implications for models of the lexicon-phonetics interface. Indeed, classical generative approaches to phonology maintain a strict division between abstract underlying representations and surface phonetic implementation, with lexical frequency having no principled role in determining phonological outcomes. Under such models, the observed frequency effect on duration would need to be attributed to post-lexical phonetic processes entirely independent of phonological structure, and the null effect on voicing would remain theoretically unmotivated.

Conversely, usage-based models (Bybee, 2001, 2007; Pierrehumbert, 2002, 2016) provide a more coherent framework for integrating these findings. In exemplar-based accounts, lexical representations consist of clouds of phonetically detailed tokens accumulated through experience, and production involves selecting from these distributions based on contextual and communicative factors. High-frequency words accumulate more tokens and undergo greater articulatory routinisation, leading to phonetic reduction, but, crucially, this reduction is gradient and context-sensitive rather than categorical. The findings presented in this work align with this framework: lexical frequency modulates fine-grained durational properties while leaving categorical voicing distinctions largely unaffected, consistent with the view that frequency operates on stored phonetic detail rather than on abstract phonological features.

Thus, the patterns observed for Italian dental affricates can be situated within a broader cross-linguistic picture that strengthens support for exemplar-based models of lexical representation. Frequency effects on phonetic realisation have been documented across typologically diverse languages, with the most robust effects consistently appearing in segmental duration. In English, numerous studies have demonstrated duration reduction in high-frequency words (Bell et al., 2009; Gahl, 2008), with additional effects on vowel quality (Munson & Solomon, 2004) and consonant realisation (Raymond et al., 2006). Similar patterns have been observed in other Germanic languages, including Dutch (Pluymaekers et al., 2005) and Romance languages such as Spanish (Torreira & Ernestus, 2011) and French (Adda-Decker et al., 2005). Furthermore, Hashimoto's (2021, 2023) research on Japanese has suggested that reduction patterns may be lexicalised rather than purely context-dependent, and these findings were interpreted within Exemplar Theory as reflecting ease of production target creation. Similar patterns have been documented in tone languages, such as Dalian Mandarin (Bi & Chen, 2022) or Taiwan Southern Min (Wang, 2022).

Therefore, the cross-linguistic consistency of frequency effects on duration has been documented across stress-timed (English, German), syllable-timed (Romance languages), and mora-timed (Japanese) prosodic systems, as well as in tone languages (Mandarin, Taiwan Southern Min). This strengthens the hypothesis that phonetic variants are stored as part of lexical representations rather than computed purely online. This lexicalisation hypothesis receives further support from the observation that informativity effects are more resistant to modulation by prosodic context than contextual surprising effects, as demonstrated particularly clearly in Wang's (2022) analysis of Taiwan Southern Min. The finding that words with generally high predictability show reduction even in locally unpredictable contexts (documented in Japanese, Mandarin, and Taiwan Southern Min) suggests that usage-based phonetic variants may constitute a universal property of the human lexical system.

Once word-level reduction is controlled for, the compensatory lengthening of affricates in high-frequency Italian words extends this cross-linguistic picture. Rather than showing straightforward reduction, Italian dental affricates exhibit relative preservation within otherwise reduced high-frequency words. This pattern aligns with listener-oriented models of speech production (Lindblom, 1990), where articulatory reduction is constrained by the need to maintain sufficient acoustic-phonetic information for successful communication. The fact that this compensatory mechanism is observed even for a phonologically marked and lexically sparse consonant class like dental affricates underscores the robustness of frequency-driven phonetic modulation across different levels of phonological structure. This could suggest that speakers actively manage phonetic variation to preserve segmental identity in contexts where overall word-level reduction might otherwise compromise intelligibility.

Finally, the present results contribute to ongoing theoretical debates regarding the phonological status of Italian dental affricates. From a classical phonological perspective, the discussion on the phonological status of Italian dental affricates has remained problematic. This is primarily due to the presence of only one minimal pair opposing voiceless and voiced variants, linked to the term *razza*: 'race/breed' with [t:s:] and '(fish) ray' with [d:z:]. In classical generative frameworks, one would expect either /ts/ or /dz/ to represent the underlying phonological form, with the surface alternation derived through context-specific phonological rules. If frequency effects were to systematically favour one variant over the other, this would provide evidence for identifying the favoured variant as underlying. However, the present findings complicate this picture: lexical frequency does not systematically determine voicing category, suggesting that neither /ts/ nor /dz/ can be straightforwardly identified as "underlying" based on frequency-conditioned pref-

erences. This creates theoretical tension within models that maintain strict separation between lexical access and phonological implementation, as it remains unclear what principle would govern the distribution of voiced and voiceless variants if not frequency-based routinisation, phonological context, or etymological convention alone.

Conversely, usage-based models (Bybee, 2001, 2007; Pierrehumbert, 2002, 2016) provide a more coherent account. Rather than assuming a single underlying form with categorical derivational rules, these frameworks treat the distribution of voiced and voiceless dental affricates as emerging from the interplay of multiple gradient factors: phonological context (which strongly conditions voicing, as demonstrated in Section 4.1), sociolinguistic variation (regional origin, speech style; see also Meluzzi, 2020), and fine-grained phonetic detail accumulated through usage. High-frequency items undergo articulatory routinisation, but this routinisation operates primarily at the level of gradient phonetic implementation (e.g., duration) rather than determining categorical phonological choices such as voicing. This explains why frequency effects are observed for affricate duration but not for voicing category: the former reflects stored phonetic detail shaped by cumulative linguistic experience, while the latter is determined by phonological structure, sociolinguistic conditioning, and contextual factors that are not straightforwardly reducible to lexical frequency alone.

The present findings, thus, support theoretical models that reject categorical phonological–phonetic distinctions in favour of gradient, usage-driven patterns where phonological 'defaults' emerge statistically from speakers' linguistic experience rather than from underlying abstract representations.

6. Conclusions and Further Perspectives

This study examines the voicing and duration of Italian dental affricates in spontaneous dialogue and read speech, addressing three research questions: whether lexical frequency predicts systematic differences in affricate realisation; how speech style shapes affricate production and interacts with lexical frequency; and whether other sociolinguistic factors contribute to explaining dental affricate variability. Our initial hypothesis was that rare and infrequent words would show greater variability, particularly in duration, and would vary more across speech styles than high-frequency words.

Our findings reveal a more nuanced picture. Indeed, lexical frequency does not predict categorical voicing distinctions once between-speaker and between-item variability are appropriately modelled, but it does exert a significant positive effect on affricate duration: high-frequency words contain relatively longer affricates, reflecting a compensatory mechanism whereby segmental preservation counterbalances word-level reduction. Speech style significantly affects duration, with reading favouring longer realisations, but frequency effects operate uniformly across styles rather than showing the context-dependent variability we had anticipated. Sociolinguistic factors, such as speakers' geographical origin and phonological context, emerged as the primary determinants of voicing, with substantial between-speaker variation indicating that Italian dental affricate voicing constitutes a sociolinguistic variable rather than being lexically conditioned.

As discussed in the previous section, these results challenge classical phonological accounts that assume frequency-driven categorical shifts and instead support usage-based models in which lexical frequency modulates gradient phonetic implementation while categorical phonological choices are determined by phonological structure and sociolinguistic conditioning. The dissociation between categorical voicing and gradient duration underscores the complexity of the lexicon-phonetics interface and demonstrates that frequency effects operate differentially across levels of linguistic representation. For Italian dental affricates specifically, the findings suggest that the distribution of voiced and voiceless variants emerges from the interplay of phonological context, regional convention, and gradient

phonetic factors, rather than from a single underlying specification or frequency-based routinisation alone.

Future studies should expand the lexical sample to include a broader range of frequency values and etymological backgrounds (within the structural limitations of Italian vocabulary), potentially revealing non-linear frequency effects that may be obscured in the current dataset. Perception experiments could test whether listeners are sensitive to the frequency-conditioned phonetic differences identified in production, providing crucial evidence for the functional relevance of these patterns. Finally, cross-linguistic investigations examining frequency effects on phonological contrasts in typologically diverse languages would provide crucial evidence for distinguishing universal cognitive principles from language-specific organisational patterns. Comparative studies could indeed confirm whether frequency consistently favours phonetically unmarked variants across different markedness hierarchies and morphological systems, ultimately contributing to more robust theoretical frameworks that capture both universal constraints and language-specific variation principles.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/languages11050087/s1, Table S1: Maptask; Table S2: Sentence reading task.

Author Contributions: The authors have jointly conceived and developed the present contribution. However, per the requirements of the Italian Academy, C.M. is responsible for methodology, and for Sections 1, 2.1, 3.1, 3.2, 4.2, and 6; N.N. is responsible for data collection and annotations, other than for writing Sections 2.2, 3.3, 4.1, and 5. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This research was not subjected to IRB approval: according to University of Milan policies, studies not involving sensitive information from marginalized social groups are exempted from IRB approval.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Acknowledgments: The authors wish to thank the reviewers and the editors for their patience and the useful suggestions; obviously, all mistakes are our responsibility. Special thanks are also due to the college students of Giasone del Maino (Pavia, Italy) for taking part in this investigation. We also wish to thank Yuka Naito for the valuable insights into the statistical analysis.

Conflicts of Interest: The authors declare no conflict of interest.

References

Adda-Decker, M., de Mareüil, P. B., Adda, G., & Lamel, L. (2005). Investigating syllabic structures and their variation in spontaneous French. *Speech Communication*, *46*(2), 119–139. [CrossRef]

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56. [CrossRef] [PubMed]

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. [CrossRef]

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111. [CrossRef]

Bellocchi, S., Bonifaci, P., & Burani, C. (2016). Lexicality, frequency and stress assignment effects in bilingual children reading Italian as a second language. *Bilingualism: Language and Cognition*, *19*(1), 89–105. [CrossRef]

Bi, Y., & Chen, Y. (2022). The effects of lexical frequency and homophone neighborhood density on incomplete tonal neutralization. *Frontiers in Psychology*, *13*, 867353. [CrossRef] [PubMed]

Bybee, J. (2001). *Phonology and language use*. Cambridge University Press.
Bybee, J. (2007). *Frequency of use and the organization of language*. Cambridge University Press.
Canepari, L. (1980). *Italiano standard e pronunce regionali*. Cluep.
Canepari, L. (1997). *Introduzione alla fonetica*. Einaudi.
Celata, C. (2004). *Acquisizione e mutamento di categorie fonologiche: Le affricate in italiano*. FrancoAngeli.
Crocco, C. (2001). I corpora AVIP e CLIPS: Il problema della codifica e della rappresentazione degli italiani regionali. In F. Fusco, & C. Marcato (Eds.), *Plurilinguismo. Contatti e culture* (pp. 151–164). Forum Editore.
Dediu, D., Lin, J., Moisik, S. R., & Moran, S. (2024). Dental fricatives: Patterning, evolution, and factors affecting a rare class of speech sounds. In F. A. Karakostis, & G. Jäger (Eds.), *Biocultural evolution: An agenda for integrative approaches* (pp. 143–178). Kerns Verlag. [CrossRef]
De Dominicis, A. (1999). *Fonologia comparata delle principali lingue europee moderne*. Clueb.
Elff, M. (2022). *mclogit: Multinomial logit models, with or without random effects or overdispersion* (R package version 0.9, 6(10.32614)). Available online: http://melff.github.io/mclogit/ (accessed on 2 April 2026).
Foulkes, P., Docherty, G., & Jones, M. (2011). Analyzing stops. In M. Di Paolo, & M. Yaeger-Dror (Eds.), *Sociophonetics: A student's guide* (pp. 58–71). Routledge.
Gahl, S. (2008). "Time" and "Thyme" are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*, 474–496. [CrossRef]
Gahl, S., & Strand, J. (2016). Many neighborhoods: Phonological and perceptual neighborhood density in lexical production and perception. *Journal of Memory and Language*, *89*, 162–178. [CrossRef]
Hashimoto, D. (2021). Probabilistic reduction and mental accumulation in Japanese: Frequency, contextual predictability, and average predictability. *Journal of Phonetics*, *87*, 101061. [CrossRef]
Hashimoto, D. (2023). The effect of verbal conjugation predictability on speech signal. *Morphology*, *33*(1), 41–63. [CrossRef]
Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, *21*(2), 233–256. [CrossRef]
Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013, July 23–26). *The TenTen corpus family*. 7th International Corpus Linguistics Conference CL (pp. 125–127), Lancaster University, Lancaster, UK.
Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson, & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). Academic Press.
Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499. [CrossRef]
Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001, May 7–11). *The effect of language model probability on pronunciation reduction*. 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 2, pp. 801–804), Salt Lake City, UT, USA.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. [CrossRef]
Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell.
Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Springer Netherlands.
Mairano, P., Nodari, R., Ardolino, F., De Iacovo, V., & Mereu, D. (2025). Inherently long consonants in contemporary Italian varieties: Regional variation and orthographic effects. *Languages*, *10*(6), 118. [CrossRef]
Martinuzzi, C., & Schertz, J. (2022). Sorry, not sorry: The independent role of multiple phonetic cues in signaling the difference between two word meanings. *Language and Speech*, *65*(1), 143–172. [CrossRef] [PubMed]
Meluzzi, C. (2016). A New Sonority Degree in the Realization of Dental Affricates /ts dz/ in Italian. In M. J. Ball, & N. Müller (Eds.), *Challenging sonority-cross-linguistic evidence* (pp. 252–275). Equinox Publishing.
Meluzzi, C. (2020). *Sociofonetica di una varietà di koinè: Le affricate dentali nell'italiano di Bolzano*. FrancoAngeli.
Meluzzi, C. (2021). Dental affricates loss and maintenance in Romance languages: A (socio)-phonetic perspective on sound change. In L. Biondi, F. Dedè, & A. Scala (Eds.), *Change in grammar: Triggers, paths, and outcomes, quaderni del sodalizio glottologico milanese* (Vol. 1, pp. 95–116). Edizioni dell'Orso.
Meunier, C., & Espesser, R. (2011). Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics*, *39*(3), 271–278. [CrossRef]
Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, *47*(5), 1048–1058. [CrossRef]
Nese, N. (2023). Variazione linguistica in un collegio universitario pavese: Uno studio in real time. In M. Castagneto, & M. Ravetto (Eds.), *La comunicazione parlata* (pp. 149–170). Aracne.

Nese, N., & Meluzzi, C. (2017). Accomodamento ed emergenza di varianti fonetiche: Le affricate dentali intermedie a Pavia e Bolzano. In C. Bertini, C. Celata, G. Lenoci, C. Meluzzi, & I. Ricci (Eds.), *Fattori sociali e biologici nella variazione fonetica/Social and biological factors in speech variation* (pp. 67–82). Officinaventuno.

Paizi, D., Burani, C., & Zoccolotti, P. (2010). List context effects in reading Italian nonwords: Can the word frequency effect be eliminated? *European Journal of Cognitive Psychology*, *22*, 1039–1065. [CrossRef]

Phillips, B. S. (2006). *Word frequency and lexical diffusion*. Palgrave McMillan.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology VII* (pp. 101–139). De Gruyter.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, *2*, 33–52. [CrossRef]

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, *118*(4), 2561–2569. [CrossRef]

Raymond, W. D., Dautricourt, R., & Hume, E. (2006). Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, *18*, 55–97. [CrossRef]

Recasens, D., & Espinosa, A. (2007). An electropalatographic and acoustic study of affricates and fricatives in two Catalan dialects. *Journal of the International Phonetic Association*, *37*, 143–172. [CrossRef]

Rossi, M., Tramutoli, L., Nese, N., Celata, C., Corona, L., & Meluzzi, C. (forthcoming). Phonetic detail in Italian homophonic simplex and complex words. In E. Bedussi, C. Celata, L. Corona, M. Frontera, C. Meluzzi, N. Nese, D. Piccardi, M. Rossi, & L.Tramutoli (Eds.), *La voce della grammatica/The sound of grammar*. Officinaventuno.

Savy, R. (1999). Riduzioni foniche nella morfologia del sintagma nominale nel parlato spontaneo: Indagine quantitativa e aspetti strutturali. In P. Benincà, L. Vanelli, & A. Mioni (Eds.), *Fonologia e morfologia dell'italiano e dei dialetti d'Italia* (pp. 1000–1021). Bulzoni.

Sbacco, L., & Meluzzi, C. (2023). Inter-speaker accommodation and within-dialect variability. Dental affricates and fricative realisation in Marchigiano. *Lingue e Linguaggi*, *56*, 345–359.

Schettino, L., & Cotugno, F. (2025). Quantifying and characterizing phonetic reduction in Italian natural speech. *Languages*, *10*(1), 14. [CrossRef]

Schmitz, D., & Baer-Henney, D. (2024, July 2–5). *Morphology renders homophonous segments phonetically different: Word-final /s/ in German*. Proceedings of Speech Prosody (pp. 587–591), Leiden, The Netherlands.

Solé, M.-J. (2015, August 10–14). *Acoustic evidence of articulatory adjustments to sustain voicing during voiced stops*. 18th International Congress of Phonetic Sciences (pp. 1–20), University of Glasgow, Glasgow, UK.

Stevens, M., & Hajek, J. (2007, August 6–10). *Towards a phonetic conspectus of preaspiration: Acoustic evidence from Sienese Italian*. 16th International Congress of Phonetic Sciences (ICPhS16) (pp. 429–432), Saarbrücken, Germany.

Stevens, M., & Hajek, J. (2010, December 14–16). *Preaspirated /pp tt kk/ in standard Italian: A sociophonetic vs. phonetic analysis*. 13th Australasian International Conference of Speech Science and Technology (pp. 1–4), Melbourne, Australia.

Tang, K., & Shaw, J. A. (2021). Prosody leaks into the memories of words. *Cognition*, *210*, 104601. [CrossRef]

Telmon, T. (2003). Varietà regionali. In A. A. Sobrero (Ed.), *Introduzione all'italiano contemporaneo* (pp. 93–149). Laterza.

Torreira, F., & Ernestus, M. (2011). Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology*, *2*(3), 331–353. [CrossRef]

van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. [CrossRef] [PubMed]

Wang, S. F. (2022). The interaction between predictability and pre-boundary lengthening on syllable duration in Taiwan Southern Min. *Phonetica*, *79*(4), 31352. [CrossRef] [PubMed]

Wheeler, M. W. (2005). *The phonology of Catalan*. Oxford University Press.

Żygis, M. (2008). On the avoidance of voiced sibilant affricates. *ZAS Papers in Linguistics*, *49*, 23–45. [CrossRef]

Żygis, M., Fuchs, S., & Koening, L. (2012). Phonetic explanations for the infrequency of voiced sibilant affricates across languages. *Laboratory Phonology*, *3*(2), 299–336. [CrossRef]