Liaison Phenomena Involving Word-Final /n/ and Word-Initial /p, t, k/ in Modern Greek: a Codification of the Observed Variation intended for Use in TTS Synthesis

Constandinos Kalimeris[1], *George Mikros*[2], *Stelios Bakamidis*[1]

[1]Institute for Language and Speech Processing,
Epidavrou & Artemidos 6, Marousi, 151 25, Athens, Greece.
[2]National & Kapodistrian University of Athens,
Panepistimioupoli, Ilissia, 157 84, Athens, Greece.
c_kal@ilsp.gr, gmikros@isll.uoa.gr, bakam@ilsp.gr

Abstract

The present paper is a preliminary attempt to codify 'liaison' (i.e. assimilation and deletion) phenomena in Modern Greek involving the word-final /n/ of certain high-frequency function words and the word-initial voiceless stop (/p/, /t/ or /k/) of the following word. In natural speech, every relevant phoneme combination affected by these post-lexical phonological processes will generate one phonetic realisation from a set of legitimate variants. The variants' distribution is not random but subject to both linguistic and extralinguistic conditions. This work also explores possible ways to exploit the findings of recent sociolinguistic research on Modern Greek with a view to accommodating the observed variation and the rules codifying it within the framework of text-to-speech (TTS) synthesis. Such a development is expected to improve synthetic speech output, both in terms of naturalness and intelligibility.

1. Introduction

Variation in speech production first received special attention by the Neogrammarian scholars between the 19th and 20th centuries, as a possible cause or vehicle of language change [1]. Quantitative sociolinguistic research conducted since the 1960's has revealed that variation in speech production is not random but subject to a number of competing factors, both socio-pragmatic (e.g. age, sex, social status, setting, formality level, etc.) and purely linguistic (e.g. phonological environment, word length, stress position, syntactic context, etc.) [2], [3]. Variation can be thought of as limited within a finite set of discrete identifiable variants. Their skillful manipulation in natural speech is the unmistakable sign of a linguistically mature native speaker [4].

Quantitative work on principled variation in speech production in Modern Greek has started relevantly recently and has focused mainly on the presence or absence of prenasalisation during the production of voiced stop sounds (symbolized by <μπ, ντ, γκ, γγ> in orthographic representations) and on the 'liaison' phenomena involving sequences of word-final /n/ and word-initial voiceless stop phonemes [5], [6], [7], [8], [9]. It has been found that the presence or absence of nasal elements in these contexts and the blocking or allowing of the relevant phonological processes are related to the code-switching process between careful and casual speaking styles. The evidence from field research suggests that speakers manipulate the available variation possibilities systematically, although not necessarily consciously [9].

The main aim of this work is to codify in the form of multiple output re-write rules the phonetic variation involving sequences of word-final /n/ and word-initial /p/, /t/ or /k/, as this has been attested in natural speech production in Modern Greek. Our secondary aim is to briefly explore fruitful ways in which these rules could be incorporated in text-to-speech (TTS) synthesis systems.

The structure of this paper is as follows: Section 2 discusses some positive effects on the quality and flexibility of synthetic speech which are expected to stem from the incorporation in TTS systems of models reproducing the variation under examination. Section 3 introduces a codification schema that can accommodate attested variant forms. Section 4 presents some morphosyntactic constraints limiting the distribution of the variation phenomena under examination. In Section 5, the proposed codification schema is used for the generation of candidate variant forms to be exploited in further TTS synthesis research. Section 6 discusses, in the context of TTS synthesis applications, the issue of optimal variant-selection based on sociolinguistic evidence and briefly outlines a relevant experiment currently under way. Section 7 concludes the paper.

2. The need to reproduce the attested variation in TTS synthesis

Realism and descriptive completeness in the theoretical analysis underlying and supporting a TTS synthesizer are desirable if one aims to develop a system intended to be used for different purposes and in a variety of contexts of interaction. If such a system develops into a commercial product, its output will have to sound sufficiently natural to pass the aesthetic test posed by its end users. Users will judge the system with their perfect, yet tacit and instinctive, *qualitative* knowledge of the communicative norms that govern spoken interaction in real life. TTS engineers attempting to improve the naturalness of their systems' output will have to do so by simulating natural speech production through the manipulation of explicitly postulated (but necessarily imperfect) *quantitative* parameters. Given the findings of recent quantitative research (see Section 1), such explicit quantitative criteria could be safely extracted only through analysis of corpora of natural speech.

Not only naturalness, however, but also intelligibility is a requirement in synthetic speech, especially when synthetic

speech systems are used by people with special abilities or in (physical or virtual) communicative settings with high levels of noise. In these cases, the requirement for intelligibility will normally override that for naturalness. Although there is no quantitative evidence yet as to the distribution of such phenomena in Modern Greek (cf. [10]), phonetic variation, in principle at least, can lead to semantic ambiguity within the limits of a syntactic phrase. In speech synthesis for special purposes, overlap between the numerous phonetic variants of different linguistic units (see below) can be kept to a minimum through artificial suppression of the potential for variation. This, while at the cost of naturalness, can be in the interests of intelligibility.

Systems can be designed to produce utterances which, though not natural-sounding in all conceivable contexts of use, will be maximally comprehensible, as the phones (the concrete sounds) that will comprise them will be maximally close to the phonemes of Modern Greek, i.e. to the sounds performing semantic distinction functions in that language. Such a system, for example, would always produce the signal [tinpíra] if fed with the input string <την πείρα> (corresponding to the linguistic unit / #tin##píra# /, "the experience"), while, at the same time, would always produce the distinct signal [tinbíra] if fed with the input string <την μπύρα> (corresponding to the linguistic unit / #tin##bíra# /, "the beer"). In natural speech, however, the phonetic variants realising the two units overlap; for example, both the signals [tibíra] and [timbíra] can be legitimately used to refer to either of the two meanings.

3. An attempt towards the formal interrelation of variant forms

Three different phonological processes (Processes 1, 2 and 3; see Table 1) may be used to account for the four types of phonetic realisation commonly attested in the data (Types (ii), (iii), (iv) and (v)) and the maximally comprehensible but rather 'artificial' type (Type (i); see Section 2).[1] One process or one combination of processes apply to each instance of post-lexical sequences of /-n#/ and /#p-/, /#t-/ or /#k-/ so as to produce the variant phonetic realisation appropriate to the communicative context at hand. When two processes are required for the production of a variant (as in cases (iii) and (iv)), these apply serially, in the given order, so that the output of the first process constitutes the input of the second [11].

The four typical realisations of the phrase / #stin##póli# /, <στην πόλη>, "to/at the city", may be used to illustrate the five variant categories of phonetic feature combinations and the phonological processes giving rise to them, according to the schema in Table 1 (where "#" represents a word boundary and "∅" the trace of a deleted phoneme). Process 0, being the 'Null Process', causes no changes to the underlying sounds. Process 1 is a 'Place Assimilation', by application of which the word-final nasal assimilates to the place of articulation of the next word's word-initial voiceless stop consonant (progressive assimilation). Process 2 is a 'Voicing Assimilation', by which the word-initial voiceless stop consonant, and the /s/ which *may* follow it, assimilate to the voicing setting of the preceding word-final nasal (regressive assimilation). Finally, Process 3 deletes the word-final nasal.[2]

Example Input	Applying Processes	Example Output	Example Realisation	Type of Variant
/n##p/	0	/n##p/	[sti**np**óli]	(i)
	1	/m##p/	[sti**mp**óli]	(ii)
	2 ➞ 1	/m##b/	[sti**mb**óli]	(iii)
	2 ➞ 3	/∅##b/	[sti**b**óli]	(iv)
	3	/∅##p/	[sti**p**óli]	(v)

Table 1: A proposed schema of post-lexical phonological processes, accounting for attested variant phonetic forms realising sequences of /-n#/ and /#p-/, /#t-/ or /#k-/.
"#": Word Boundary. "∅": Trace of a deleted phoneme.
"0": Null process. "1": regressive Place Assimilation.
"2": progressive Voicing Assimilation.
"3": Deletion of the word-final nasal.

It must be stressed here that none of the commonly attested types of phonetic realisation (Types (ii) – (v)) constitutes an 'accent' in its own right. A person invariantly realising all relevant post-lexical sequences of phonemes according to the same type of variant would certainly sound unnatural; in fact, much of the 'unnaturalness' of TTS systems may be attributed to the same flaw. *Variation is inherent in speech production*. All variant forms are present in the linguistic repertoire of any one individual, albeit in different numbers. Furthermore, no two individuals display the same pattern of frequencies for the four variant types of realisation. Sociolinguistic research, however, commonly invokes a standardized pattern of frequencies as a sort of 'mean' for groups of people identified by the same set of social characteristics, such as common age, gender, class, etc. [2], [3], [9] (see Section 6).

Invariability would be justified only when maximum intelligibility were required. In that case, Type (i) realisations would probably be the safest bet: being rather 'artificial', they can be distributed evenly within utterances.

4. Morphosyntactic constraints on the input of the phonological processes

A TTS system implementing (some form of) the codification system presented in this paper will, at some point, have to determine which pairs of words in its input are relevant to the phonological processes presented in Section 3. This is a task *distinct from and temporally preceding* the task of deciding which of the pronunciation options (i) to (v) offered by the schema of Table 1 will eventually be used for the coarticulation of each legitimate pair of input words.

At this point of our research it appears that the first word (or "W1") of each legitimate pair can only be a function word. Examining the list of the 1,000 most frequent orthographic words in the Hellenic National Corpus (HNC), the online electronic text corpus developed by the Institute for Language and Speech Processing (ILSP) ([12], [13]), we have identified only 13 words that can undoubtedly function as W1 and, therefore, trigger the rules. All of them are function words: the forms of the definite article <τον, την, των>, the form <έναν> of the indefinite article, the negative particles <δεν, μην>, the inflected prepositions <στον, στην, στων>

and the conjunctions <σαν, αν, όταν, πριν> (<τον, την> can also represent the weak forms of the personal pronoun).

Despite their small number, the above word types represent in total 1,975,981 orthographic word tokens in HNC, or 5.78% of the 34,158,816 orthographic strings comprising the corpus. The high frequency of these types indicates that a successful tackling by a TTS system of the 'liaison' phenomena we are dealing with here can greatly contribute to the improvement of its performance.

The previous morphosyntactic constraint on W1 needs to be mirrored by one on W2: the second word of each co-articulated pair will be a content word, belonging, in the majority of cases, to one of the four prototypical inflected word categories of Greek: a noun or adjective will normally follow <την, τον, των, στην, στον, στων, έναν, σαν>, while <δεν, μην> will normally be followed by a verb or pronoun. <την, τον>, when representing a pronoun, can also be followed by a verb.

At present, our research does not extend to cases when both W1 and W2 are content words. Further work is required so that our proposed schema of phonological processes is adequately modified to cover such cases as well.

5. A set of Multiple-Output re-write rules

The schema of phonological processes outlined in Table 1 can also be used inversely, i.e. not only for the accommodation of attested forms, but also for the production of candidate variant phonetic forms in cases when /n/ is followed by voiceless stops other than /p/. Inspired by recent work in other fields [14], we have used our schema to generate all possible phonetic variants for post-lexical combinations involving /-n#/ and each of the segments generally accepted as the allophones of the stop phonemes of Modern Greek [15], [16] (see Table 2). "†" marks instances of redundant application of a (set of) process(es), while "^" is the "*not*" logical operator: "p^(s)" means "/p/ followed by any segment other than /s/".

MO Rule	Input string		Variant Output string				
			i	ii	iii	iv	v
1	n##p^(s)	→	np	mp	mb	b	p
2	n##t^(s)	→	nt	nt†	nd	d	t
3	n##c	→	nc	ɲc	ɲɟ	ɟ	c
4	n##k^(s)	→	nk	ŋk	ŋg	g	k
5	n##ps	→	nps	mps	mbz	bz	ps
6	n##ts	→	nts	nts†	ndz	dz	ts
7	n##ks	→	nks	ŋks	ŋgz	gz	ks

Table 2: Phonetic variants for post-lexical combinations of /-n#/ + voiceless stop segment, as generated by the schema of Table 1. "MO": multiple-output. "^": NOT logical operator. "†": processes applying redundantly.

The schema creates overlapping forms only in the case of post-lexical sequences involving dental stops (/t/): in M(ultiple) O(utput) Rules 2 and 6, Type (ii) variants arise from the redundant application of Process 1 on the underlying sequence /-n##t-/. Note that the attested variant types [mbz], [bz], [ndz], [dz], [ŋgz] and [gz] (MO Rules 5 – 7, variant Types (iii) and (iv)) cannot be accounted for unless Process 2 makes additional provision (as it does) for the assimilation of "the /s/ which may follow" the word-initial voiceless stop (see Section 3). The very existence of these variants could constitute an argument in favour of a monophonematic interpretation of the sounds commonly represented as [ps], [ts] and [ks] ([10], [15]).

For a TTS application incorporating our codification schema of the 'liaison' phenomena under examination, Table 2 merely represents a starting point. The multiple output of the MO Rules presented in this section can and, in all likelihood, will have to be modified after further relevant experimentation. For example, certain generated variants may prove redundant because they may prove indistinguishable from other variants in auditory terms. In such cases, a merging of variant types may prove necessary. Also, the schema of phonological processes presented in Table 1 can be transformed into a strictly ordered system able to function as a generation matrix. This can be used for the production of more candidate variant phonetic forms, involving word-final /n/ and, not just voiceless stops, but also all the remaining consonant phonemes of Greek. Since the definitions of the different natural classes of sounds require different degrees of descriptive complexity, the matrix's output will certainly be in need of pruning so that, not only redundant, but also ungrammatical generated forms are eliminated.

6. Controlling the rules' output: an experimental approach

A question naturally arising from the discussion so far is how one could practically exploit a system of categories like the one outlined in Sections 3 and 5 so as to produce synthetic speech of improved quality in terms of naturalness. An algorithm can be devised to enable optimal selection of one variant every time a TTS system encounters in its input one of the post-lexical phoneme sequences listed in Table 2. It follows (see Sections 1 and 4) that the algorithm's selection criteria can, and normally will be, both socio-pragmatic and purely linguistic.

The complexity of the interrelated criteria that inform variant choice in natural speech could be artificially simplified for the purposes of synthetic speech production. A project, for example, may seek to synthetically reproduce the different speech styles of a particular individual, or a particular style, say the relaxed vernacular style, of the inhabitants of a major urban center, irrespective of speakers' social class, age, or even gender. An experiment currently under way is doing so for the sample of speakers providing the speech data in a recent field study [9]. The experiment utilizes the study's findings regarding the distribution patterns for the phonetic variants of post-lexical sequences of /n/ and /p, t, k/ in the speech of the sample of speakers as a whole. The study examined the correlation of the distribution of the 4 attested variants with 41 independent linguistic variables; by means of statistical methods it was discovered that the distribution in question correlates significantly with 8 of them. In the experiment, the relevant data are utilized by an optimal selection algorithm using probabilistic criteria. An existing TTS application is being modified to incorporate the algorithm. The algorithm will utilize pre-processing tools (such as a part-of-speech tagger and a syllabifier) developed by ILSP, and will be integrated into the grapheme-to-phoneme module of the system. The system's performance will be evaluated for naturalness by a random sample of non-

specialist individuals. The experiment and its results will be reported in a forthcoming paper.

7. Summary

This paper presented a schema for the codification of the phonetic variation involving the word-final /n/ of certain high-frequency function words and the word-initial voiceless stop phoneme of the following word. The variation was modeled as an unordered, but potentially modifiable, system of phonological processes (two different types of assimilation and a deletion), altering the underlying forms of words and giving rise to multiple surface forms. The input to the relevant re-write rules was constrained by reference to morphosyntactic criteria. The practical gains of a possible reproduction of the attested variation within the framework of TTS synthesis were discussed. The problem of controlling the multiple output of the rules in TTS synthesis by use of linguistic and extra-linguistic criteria was posed. Finally, an experimental effort to incorporate our proposed set of multiple-output rules into an existing TTS system with a view to accomplishing perceptible levels of improvement was schematically outlined.

---

[1] Although Type (i) realisations (i.e. nasals not assimilated to the place of articulation of the following voiceless stop) are attested extremely rarely, they certainly are not ungrammatical. They are commonly referred to as 'orthographic pronunciations' and are normally expected to occur in emphatic utterances. No instances of Type (i) realisations occur in the speech data we are using [9].

[2] Despite linguistic intuitions as to the contrary, Type (v) realisations are well-documented in the speech data we are using [9].

8. References

[1] Weinreich, U., Labov, W., and Herzog, M., "Empirical foundations for a theory of language change", in Lehmann, W. P. and Malkiel, Y. (eds.), *Directions for Historical Linguistics: A Symposium*, Austin and London, University of Texas Press, 95-195, 1968

[2] Labov, W., *Principles of Linguistic Change. Vol.1: Internal Factors*, Oxford, Blackwell, 1994.

[3] Labov, W., *Principles of Linguistic Change. Vol.2: Social Factors*, Oxford, Blackwell, 2001.

[4] Hymes, D., "On Communicative Competence", in Pride, J. B. and Holmes, J. (eds.), *Sociolinguistics: Selected Readings*, London, Penguin, 269-293, 1972.

[5] Arvaniti, A., "Socioliguistic patterns of prenasalization in Greek", *Proceedings of the 14th Annual Meeting of the Linguistics Department of the University of Thessaloniki*, Thessaloniki, 209-220, 1995.

[6] Arvaniti, A. and Joseph, B., "Variation in Voiced Stop Prenasalization in Greek", *Glossologia*, *Vol. 11-12*, Athens, Leader Books, 131-166, 2000.

[7] Charalambopoulos, A., Arapopoulou, M., Kokolakis, A. and Kiratzis, A., "Phonological variation: voicing – prenasalisation", *Proceedings of the 13th Annual Meeting of the Linguistics Department of the University of Thessaloniki*, Thessaloniki, 289-303, 1992. In Greek.

[8] Mikros, G., "Radio news and phonetic variation in Modern Greek", *Proceedings of the 2nd International Conference on Greek Linguistics, Vol. I*, 35-44, 1997.

[9] Mikros, G., "A Sociolinguistic Approach towards Phonological Problems of Modern Greek. Phonetic Variation in Nasal Consonants", unpublished doctoral dissertation, Department of Linguistics, University of Athens, 1999. In Greek.

[10] Kalimeris, C., "A study for the extraction of phonological information from the Hellenic National Corpus (HNC) with a view to producing Minimal Pairs and computing Functional Loads for Modern Greek", unpublished master's thesis, University of Athens (Department of Linguistics), National Technical University of Athens (Department of Computer Engineering and Informatics) and Institute for Language and Speech Processing (Athens), 2004. In Greek.

[11] Spencer, A., *Phonology: Theory and Description*, Oxford, Blackwell Publishers, 1996.

[12] Hatzigeorgiu, N., Gavrilidou, M., Piperidis, S., Carayannis, G., Papakostopoulou, A., Spiliotopoulou, A., Vacalopoulou, A., Labropoulou, P., Mantzari, E., Papageorgiou, H. and Demiros, I., "Design and implementation of the online ILSP Greek Corpus", *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, 1737-1742, 2000. (HNC on the web at http://hnc.ilsp.gr/en/)

[13] Hatzigeorgiu, N., Mikros, G. and Carayannis, G., "Word length, word frequencies and Zipf's law in the Greek language", *Journal of Quantitative Linguistics, Vol. 8*, 175-185, 2001.

[14] Bernsen, N. O., "Multimodality in Language and Speech systems: from Theory to Design Support Tool", in Granström, B., House, D. and Karlsson, I., (Eds.), *Multimodality in Language and Speech Systems*, Dordrecht, Kluwer, 93-149, 2002.

[15] Kotropoulos, K., Mavromatidou, P. and Pitas, I., "Phones and phonemes of Modern Greek", *Proceedings of the 21th Annual Meeting of the Linguistics Department of the University of Thessaloniki*, Thessaloniki, 2000. In Greek.

[16] Bakamidis, S., and Carayannis, G., "PHONEMIA: a phoneme transcription system for speech synthesis in Modern Greek", *Speech Communication 6,2*, 159-170, 1987.