Frontema: A Unit for the Study of Word Boundaries

José Albos Rodrigues, Benedito Guimarães Aguiar Neto** and Maria Auxiliadora Bezerra****

*Computer and Systems Department – Federal University of Campina Grande
albos@dsc.ufcg.edu.br
** Department of Electrical Engineering – Federal University of Campina Grande
bganeto@dee.ufcg.edu.br
*** Department of Letters – Federal University of Campina Grande
cidabezerra@uol.com.br

Abstract

This work presents the *frontema* as a unit created to facilitate the study of phenomena regarding co-articulatory effects in word boundaries. *Frontema* is the orthographic representation of written language as well as of the acoustic unit corresponding to word boundaries that conventionally form a written text. The concept of *frontema*, as well as its materialization, simplify the process of discovery of co-articulatory rules, thus contributing to facilitate the prosodic modeling of phrases and, so, reducing the complexity of the process of automatic generation of prosody in systems of speech synthesis by concatenating acoustic units previously recorded.

1. Introduction

Research on prosody aiming to develop systems of speech synthesis by the concatenating method highlights the need for further studies to make viable the prosodic modeling of effects caused by co-articulatory phenomena that happen during phonation. Such effects are also present when reading an orthographically written text, as a result of the reconfiguration of the articulatory parts of the speech organs, while passing from the phonation of one syllable (or word) to another. The effects mentioned above occur both between syllables of the same word and between adjacent words, due to the fact that the punctuation and spaces between such words, expressed on writing, are not always taken into account by readers while reading a text. Such a fact makes the speech unit be different from the written unit, meaning that the prosodic modeling cannot be made taking into account only the written language, but the spoken language as well. Moreover, the language knowledge must refer to a specific speaker or to a group of speakers that share the same co-articulatory habits.

Research in this area is motivated by the need to know more detail about co-articulatory phenomena, in order to develop prosodic models that verify, on synthetic speech, an acceptable naturalness. Such naturalness is expected in text-speech translator, in order for them to be accepted by potential users that expect that the speech produced by such translators have close resemblance with the speech from reading a text made by a human being, mainly when it is a long text.

Although there are considerable advances in the development of prosodic models to words [1], the prosodic modeling of phrases still requires improvement, especially regarding co-articulatory effects in word boundaries, where a significant amount of co-articulatory phenomena is concentrated [2]. The existence of several linguistic variations in Brazilian Portuguese, linked to the fact that the presence of co-articulatory phenomena in word boundaries is frequent, demands the development of exploratory research, aiming to know a higher number of possibilities of co-articulatory phenomena, in order to obtain effective prosodic models.

The process of spoken text translation starts, in general, with a text analysis developed in light of written language rules in order to extract necessary information for the speech synthesis *per se*. However, the co-articulatory rules are executed after the text analysis, using mechanisms that send the rules imposed by the dynamics of the articulatory organs that, generally, do not follow the representation of prosodic characteristics orthographically expressed in texts rigorously, especially the intrinsic prosody of words. This means that letters and diacritics, used for representing prosodic characteristics, are not usually taken into account during phonation while reading a text. Thus, a spoken text translator must include both the writing and the speech rules. Therefore, in order to acquire naturalness in a spoken text conversion process by artificial means, it is necessary to take writing rules during the text analysis into account, and, then, "violate" or ignore the same rules during the synthesis process, substituting them for prosodic rules commonly practiced by a speaker or group of speakers. In other words, in order to imitate the human spoken language corresponding to a text, it is necessary to go against writing rules.

Several pieces of research have been carried out in order to understand the dimension of the impact of such phenomena better. It has been verified that, in fact, word boundaries represent a broad field for research [3]. At UFCG (the institution where we work), studies on word boundaries were started systematizing the universe of language possibilities, especially related to the co-articulatory effects. In this line of thought, a study of characterization of boundaries was developed, observing phoneme border [4]. However, the need for a language and a methodology for the specification of word border aspects was noticed, following the perspective of written and spoken language simultaneously, bearing in mind that the orthographic representation differs, in several cases, from the phonetic and phonological representations. Such a case happens due to the fact that there is no bi-univocal correspondence among letter, phone and phoneme. In order to overcome such a shortcoming, it was decided to develop a study on boundaries in view of characterizing it, using a single language for representation of the corresponding written and spoken language, opting for the use of orthography, for the text is already represented by it. So, it was decided to represent speech with the same linguistic elements used in written language. Such a decision made the visibility more realistic regarding the difference between written and spoken language, even making the study of different speakers possible.

It is important to highlight the fact that, when a co-articulatory phenomenon occurs in word boundaries, the result is generally a word that does not make part of the lexicon of a language, which is constituted of syllables that can or cannot be a conventional syllable of that language. Moreover, co-articulatory effects produce segments that, if expressed using written language, do not belong to the group of possibilities of the language, as for example, four vowels together, more than one stressed syllable in the same word, among others. Because of this fact, it was decided to consider such segments as a new unit that, for not being known, was called *frontema*, and it will be described next.

2. The concept of *frontema*

The term *frontema* was created in order to denote the segment of written language and its corresponding spoken language referring to word boundaries of a text. It was decided to use it in similarity with common linguistic units, as, for example, *grafema, fonema, lexema, sema*. It was used "*front*" from "frontier" [*fronteira*] and it was added "*ema*", turning out "*frontema*".

Frontema is the orthographic representation of the text corresponding to word boundaries and to its corresponding acoustic unit, referring to a speaker or group of speakers. Thus, the *frontema* represents both what is written and what is spoken, orthographically expressing a co-articulatory phenomenon present in phonation, which makes punctuation and spaces between words present in written language disappear. In this way, the *frontema* represents what is expressed in the text and what is expressed in spoken language, according to some speech habits of a determined speaker (or group of speakers). As the *frontema* has to contain information about written and spoken language, it was decided to use the following minimum notation for it:

(*X, Y*),

where ***X*** stands for the orthographic representation of border syllables in their original form, and ***Y*** stands for the orthographic representation of the corresponding speech of ***X*** (border). The pieces of information that compose the *frontema* are in between parenthesis. When ***Y*** reflects ***X***, the information of the *frontema* corresponding to ***Y*** can be omitted, and the *frontema* is marked by **(*X*)**. In such cases, the *frontema* contains only one piece of information: the written one. In the text "*gato armado*[1]", the syllables "*to ar*" correspond to border syllables "*toar*", or "*tar*", corresponding to spoken language. If there are co-articulatory phenomena, there will be one of two forms: simple joining of sounds or omitted joining of sounds. In the last form, first there is omission, and, then, joining. In the example presented above, the *frontema* will be **(*to ar*, *toar/tar*)**. There are three possibilities for spoken language: the existence of no co-articulatory phenomenon – in this case, the *frontema* would be **(*X*)** – **(*to ar*)**; the existence of simple joining of sounds **(*toar*)**; or the omission followed by the joining sounds **(*tar*)**. If there is more than one ***Y*** possibility, they will be separated by bars (/). In other words, the speaker can practice either alternative. This fact shows that the *frontema* is adaptable in order to consider linguistic variations.

Next, I present some illustrations with their respective *frontemas*:

(1) casa amarela→ (sa a, saa/sa)→casamarela
(2) este auditório→(te au, teau/tau)→esteauditório/estauditório
(3) esse aumento→(-se au, -seau/-sau)→esseaumento/essaumento
(4) quando esse→ (do e, doe/de)→quandoesse/quandesse

In all the cases above, the co-articulatory phenomenon creates a word that does not belong to the Portuguese lexicon. It is also known that what is spoken differs from what is written, and that**,** in all the four cases, there is some change in tonicity in the first word of each group. This suggests that an adequate stress is made in order to reflect the new prosodic reality after the co-articulatory phenomenon. In this case, check the following examples:

(1a) casa amarela→(sa a, saa/s → casaamarela/casámarela
(2a) este auditório→(te au, teáu/táu) → esteáuditório/estáuditório
(3a) esse aumento→(-se au, -seáu/-sáu) → esseáumento/essáumento
(4a) quando esse→(do e, doê/dê) → quandoêsse/quandêsse

3. Justifying the use of *frontema*

The differences between orthographic, phonetic and phonological representations increase the complexity of prosodic modeling of word boundaries considerably, and, consequently, the complexity of automatic generated algorithm of prosody in spoken text translators by concatenation of acoustic units previously recorded. So, it is suggested that the representation of both the text and the corresponding spoken acoustic unit be made using the same language. In this line of thought, the *frontema* can be the work unit that fills the gaps appropriately, especially because the representation of the acoustic unit makes the preservation of the intrinsic prosody possible**;** it expresses, orthographically, the introduction also by means of second stress, and, mainly, of punctuation and accent signs that come up due to co-articulatory effects. Furthermore, several other reasons justify the use of the *frontema* as an alternative to be useful as well as efficient to processes of prosodic modeling and implementation of automatic generators of prosody, as you can see in the next cases:

- The co-articulatory effects originate the acoustic units that are characterized by syllables not expressed in the texts, and, in some cases, do not make part of the group of syllabic possibilities of the language;
- The co-articulatory effects in word boundaries are significant factors determining speech naturalness;
- The existence of a huge number of speakers that use a certain co-articulatory effect is common;
- The co-articulatory effects depend on several factors, such as speaking speed, the co-articulatory context, the speaker, linguistic variation, and they can all be expressed orthographically;
- The paucity of reports about studies on word boundaries of Brazilian Portuguese words related to prosodic modeling;
- There is a huge amount of cases in which the co-articulatory effects affect the intrinsic prosody of syllables and words involved, as well as of words around them (context);

[1] For phonetic reasons, examples will be kept in Portuguese in order to avoid any inappropriate changes in meaning.

• The written language has resources to express prosodic characteristics, such as punctuation and graphic accents that can be used to remark co-articulatory effects related to prosody;
• When a co-articulatory phenomenon occurs, it is hard to locate or separate, on the acoustic wave, the corresponding sounds from the letter involved in the *frontema* (frontier).

Another important motivation for the use of *frontema* is the fact that, in the process of development of spoken-text conversion, algorithms are necessary for carrying out several analyses in the text in order to extract from the orthographic representation relevant information for the synthesis process *per se*. One of such algorithms is the prosodic model that makes the phase of automatic generation of prosody possible, during which the behaviors that the system (artificial) must present to imitate the human voice are specified. Such a process demands context analysis, using segments different in size, such as letters, syllables or words. As the letter-phoneme conversion makes the efficient recovery of close contexts impossible, it is not possible to have an efficient automatic generation of prosody either. The *frontema* facilitates several studies related to prosodic modeling and facilitates the implementation of automatic generators of prosody in spoken-text conversion.

4. Types of *frontemas*

The *frontema* can be classified according to different effects caused by co-articulatory phenomena. Next I present the classification according to the following aspects:
a) Omission or non-omission of letters;
b) Change or no change of intrinsic prosody;
c) Number of vowels in the *frontema*, resulting from the co-articulation.

In the next subsections, there is a resume for each one of these *frontemas*.

4.1 Omission of Letters

Analizing all the possibilities of language, it was observed that it is possible to have *frontemas* that can be characterized by the omission or non-omission of letters of the first word, when pronouncing the words (or syllables) one after the other. The *frontemas* without omission preserve the same letters present in written language, and the *frontemas* with omission have letters subtracted by the phonation process. In general, there is the omission of one single letter, being possible, however, to omit more than one letter, one syllable, or even one word. In Table 4.1, there are examples of each type presented:

Words	Representation of writing	Representation of speaking
(5) buraco amarelo	*co a*	*coa* or *ca*
(6) amigo abnegado	*go ab*	*goab* or *gab*
(7) fofo armado	*fo ar*	*foar* or *far*
(8) dado ignorável	*do ig*	*doig* or *dig*
(9) acabo ignorando	*bo ig*	*boig* or *big*
(10) sonho idiota	*nho i*	*nhoi* or *nhi*
(11) metrô arcaico	*trô ar*	*troar*

Table 4.1 - Examples of *frontemas* with and without omission

Table 4.1 demonstrates that in (11) it is possible not to have joining sounds. However, if there are joining sounds, there will never be omission, because the end syllable "*trô*" is tonic. On the other hand, when there are joining sounds in the other cases, from (5) to (10), such a joining can happen with or without the omission of letters, depending on the speaker, the situation or on the communicative purpose.

4.2 Prosodic Change

Such a classification of *frontema* takes into account the co-articulatory phenomena that clearly cause perceptible changes in prosodic characteristics of syllable boundaries, maybe also affecting the prosody of other syllable around (adjacent) the syllable boundaries (or not). So, there can be *frontemas* that preserve or not the intrinsic prosody of the involved syllables (or words). In Table 4.2 some examples of *frontemas* that preserve and some *frontemas* that do not preserve the original prosody of words are presented:

Words	Representation of writing	Frontema of speaking
(12) cajá ácido	*já á*	*jaá*
(13) patê ácido	*tê á*	*teá*
(14) está áspero	*tá ás*	*taás*
(15) descobri afta	*bri af*	*briáf*
(16) desci ontem	*ci on*	*ciôn*

Table 4.2 Examples of *frontemas* with and without prosodic change.

It is known that there is prosodic change in all the cases, from (12) to (16), because the last syllable of the first word in all the examples given is tonic. However, when speaking, such syllables are no longer tonic, or they remain tonic, but with a change in tonicity (reduced, in this case), becoming a secondary stress.

4.3 Vowels together

Such a classification takes into account the number of adjacent vowels resulted from the co-articulatory phenomenon, when there is joining of sounds (simple or with omission). When there is a border, such as V-V, it is possible to have vowels together, including two, three or four vowels. Therefore, when these vowels come together with border syllables, there can be *frontemas* with several vowels that are pronounced in one single aspiration. Such *frontemas* present vowels together that, according to the number of vowels, can be classified as[1]:

a) *frontema divocálico* – when there are two vowels after the co-articulation;
b) *frontema trivocálico* – when there are three vowels together;
c) *frontema quadrivocálico* – when there are four vowels together;
d) *frontema quintivocálico* – when there are five vowels together.

Next, observe Table 4.3 illustrating examples of each classification of *frontema* listed above:

[1] As the classification above was created by the author, it will be kept in its original language, Portuguese.

Words	Representation of writing	Representation of speaking
(17) parece útil	*ce ú*	*céu*
(18) pobre uaca	*bre ua*	*breuá*
(19) paraguai a	*guai a*	*guaiá*
(20) boi amarelo	*boi a*	*bóia*
(21) sabe uiraúna	*be ui*	*bui*
(22) recebi a	*bi a*	*biá*
(23) falei autorizando	*lei au*	*leiau*

Table 4.3 Examples of *frontemas* with different number of vowels together.

In (17), (21) and (22), there are *frontemas divocálicos* (*eú, ui* and *iá* respectively). In (18) and (20), are the *frontemas trivocálico* (euá and *oia*), and in (19), there is *frontema quadrivocálico (uaiá).*

5. Possibilities for the accurrence of *frontemas* in Brazilian Portuguese

An inventory was made to identify the possible word boundaries in Brazilian Portuguese, analyzing the lexicon in a dictionary [6], without taking into account verbal conjugation. Next, the possible beginning and ending syllables that can be form word boundaries were counted, eliminating repetitions, and obtaining all possible values of X, i.e., the orthographic representation of all possible word boundaries. Finally, it was asked two speakers to read each pair of words in order to check the corresponding values of Y after listening to their reading. The values of Y are the orthographic representation of the corresponding speech of X (written language). Such an experimeent showed that most of the cases represent *frontemas* without omission. Secondly, there are the *frontemas* with omission of letters. It was also possible to see that most of the cases not produce co-articulatory phenomenon. Such a fact shows that the universe of possibilities of co-articulatory effects of a language is huge.

6. Conclusion

This work highlighted that the synthesis of speech by the concatenative method, taking the syllable as the unit of work, is only acceptably natural when using a dictionary of accustic units corresponding to the conventional syllables of the language under investigation. The use of such a dictionary will offer more naturalness to the synthetic speech.

In conclusion, it is possible to say that the *frontema* is a unit of work that facilitates the compared perception between what is spoken and what is written. The study of the *frontema* will offer relevant contribution to the advances of the prosodic modeling of phrases, and, consequently, it will serve as basis for other investigations about co-articulation, mainly the ones aiming to verify the naturalness of synthetic speech.

7. References

[1] Costa Neto M. L. da, 2004. Um modelo para geração de prosódia de palavras em conversores texto-fala para a língua portuguesa falada no Brasil. Tese de Doutorado, Departamento de Engenharia Elétrica, Universidade Federal de Campina Grande.

[2] Rodrigues J. A., 2006. Um estudo sobre prosódia de fronteiras de palavras no Português Brasileiro aplicado à síntese da fala. Exame de Qualificação, Departamento de Engenharia Elétrica, Universidade Federal de Campina Grande.

[3] Rodrigues J. A.; Bezerra M. A. ; Aguiar Neto B. G., 2005. Fronteiras de palavras: um vasto campo de pesquisa sobre prosódia. In *V Encontro Nacional de Língua Falada e Escrita*, Maceió:Brasil, 2005, 128.

[4] Rodrigues J. A.; Bezerra M. A. ; Aguiar Neto B. G., 2006. Caracterização de efeitos coarticulatórios do Português Brasileiro analisando os fonemas de fronteira de palavras. In *IX Congresso Nacional de Fonética e Fonologia* e *III Congresso Internacional de Fonética e Fonologia*, Belo Horizonte: Brasil, 21.