Language as a complex system: the case of phonetic variability

Philippe Blache, Christine Meunier

Language as a complex system: the case of phonetic variability

Philippe Blache & Christine Meunier

LPL-CNRS, Université de Provence
Aix-en-Provence, France
{blache,meunier}@lpl.univ-aix.fr

Abstract

Modern linguistic theories try to give an exhaustive explanation of how language works. In this perspective, each linguistic domain, such as phonetics, phonology, syntax, pragmatics, etc., is described by means of a set of rules or properties (in other words, a grammar). However, it is a long time linguists have observed that it is not possible to give a precise description of a real life utterance within a unique domain. We illustrate this problem with the case of phonetic variability and show how different domains interact. We propose then a two-level architecture in which domain interaction is implemented by means of constraints.

1. Introduction

Recent evolution of linguistics (see for example [1] or [2]) leads to take into consideration the fact that language constitutes a complex system. First, and this point is of particular importance in cognitive linguistics, language does not constitute an autonomous cognitive process. Many recent neurolinguistics experiments seem to confirm this aspect, in particular by means of imaging techniques. In the same direction, and even more interestingly, modern linguistics stipulates now that linguistic domains such as phonetic, phonology, syntax, semantics, etc. are not autonomous modules and cannot be described separately. More precisely, each domain can have its own structuration, but linguistics has to describe their interaction as well.

Our position is situated in this perspective: linguistic information is disseminated into the different domains. Each one contains a more or less important part of information vehiculated in the message or the utterance. But at an upper level, domain interaction also produces information: emphasis for example can result from a conjunction of prosodic and syntactic information. In the same way, morphologic, semantic and pragmatic domains often interact in the interpretation of some phenomenon such as politeness. In this perspective, we think that information has to be considered as coming both from each linguistic domain but also from domain interaction which constitutes in itself a meta-level. This approach relies then on a collaborative conception of domains relations more than a dependency hierarchization between them. One consequence is that in this approach, constraints controlling the construction of a given domain do not depend directly from another domain, but from this meta-level domain interaction.

We propose in this paper to describe such a conception of linguistic information flow. In the first section, we illustrate this organization with the example of phonetic variability. We consider such phenomenon as a typical example of such an interaction: we will show how the quantity of information produced by domain interaction has direct consequence on phoneme production. In the second section, we propose a general framework explaining this two-level architecture.

2. Phonetic variability

2.1. An evidence from physical observations

In a study related to the acoustic realizations of vowels [3], we highlighted a very strong variability in the production of the vowel /a/. We compared acoustic-phonetic variability in three different types of speech: spontaneous, continuous read speech and isolated read words. Analyses of vowels extracted from these samples revealed greater dispersion of F1 and F2 values for /a/, especially in the context of spontaneous speech (see figure 1 & 2). Results obtained from a transcription task show both poor identification scores for /a/, and a correlation between subjects' responses and the acoustic properties of the items.

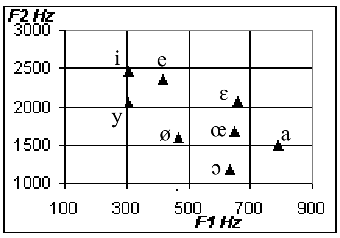


Figure 1: prototypical values of F1/F2 French vowels.

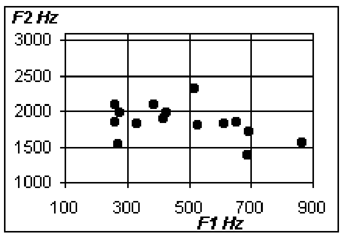


Figure 2: F1/F2 values of vowel /a/ extracted from spontaneous speech.

The importance of the variation observed in this study leads us to wonder about the need for processing of acoustical-phonetic detail in speech. A later study [4] highlighted the correct identification of the vowels /a/ once the syllabic, then lexical, context was added. How can one explain the importance of this phonetic "inaccuracy"?

2.2. Looking for Invariance

Phonetics, like other disciplines, took advantage since the beginning of the XXth century of the technical projections allowing to observe the physical characteristics of the sounds with an increasingly large precision. These technical tools probably gave the illusion that it was sufficient to physically describe speech sounds 1/ to understand the nature of the sounds of language 2/ to solve the first stage of the speech comprehension process: acoustical-phonetic decoding. From this point of view, and in order to be able to give a relevant acoustic description, "variability" was to be eliminated. Speech research, in that time, was concentrated on the characterization of invariant acoustic and/or articulatory features which made it possible to the listeners to correctly identify the sounds of the language [5] [6].

However, multiple possible realizations of the same sound unit, therefore its variability, make acoustic characterization of speech sound categories very difficult. Thus, if the listeners of a language do approximatively agree about which are the sounds present in their language, it seems nevertheless impossible to give a detail and unambiguous physical description of sound categories. Now, everyone admits that there are no acoustic invariants. It is vain thus to seek only in the speech signal these features which would enable us to identify the sounds of our language.

2.3. Different types of variability

For a long time, variability was regarded as noise making obstacle to the identification of the sounds of language. This is a complex question because there is confusion about the different phenomena which one calls variability. Indeed, it is common to call “variation” phenomena as distinct as coarticulation, speech style or random variation. Confusion is due to the fact that one is unaware of what should be a realization with no variation. Thus, is coarticulation a special type of variation compared to prototypes of isolated phonemes? Or does it form an integral part of the prototypes of speech production?

We think that it is possible to distinguish two types of variability which are, on the one hand, one that makes communication more difficult and, on the other hand, one that helps communication. In the "Theory of Information" [7], a principle stipulates that any channel contains sources of noise making obstacle to the signal transmition. Thus, when one wants to counterbalance this obstacle and thus facilitate communication, it is enough either to decrease the noise or to increase the signal. Accordingly, and with regard to speech production, we make the assumption that when the various linguistic fields are informative, the phonetic realizations can be inaccurate, or less controlled.

2.4. Phonetic realizations and linguistic information

In order to explain variability in speech communication, Lindblöm [8] proposes an adaptive model of production (Hypo & and Hyper Speech Theory). Speakers permanently adapt their production according to the situations of communication. The articulation is thus careful, if the situation makes it necessary, and less careful if that is possible without harming comprehension. Many works highlighted the existence of “top-down” processes. Ganong [9] shows that the presence of an ambiguous phonetic segment is not percieved once the lexical context is sufficient for word identification. In the same way, Warren [10] shows that subjects do not note the presence of a noise replacing a phoneme when lexical information is sufficiently non ambiguous. These results clearly show that the lack of phonetic information is not systematically an obstacle for comprehension insofar as other types of information (lexico-semantic, syntactic, pragmatic, etc) are present and constitute another “set of information”. Conversely, some accurate phonetic features can be essential for speech comprehension when they represent the only elements of clarification.

In this way, we hypothesize that the realizations of speech sounds depend on the quantity of information brought by other linguistic fields. Our assumption is that there is a relation between the quantity of linguistic information and the dispersion of the phonetic realizations: if linguistic information is less conspicuous , then phonetic realization will be closer to the prototypical form.

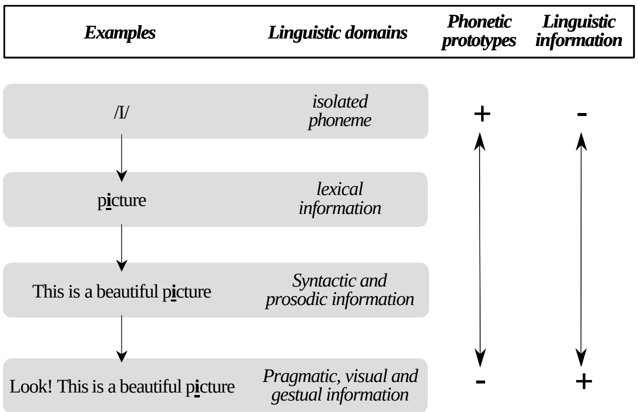


Figure 3 : phonetic variation according to linguistic information

Conversely, if contextual information is present, dispersion of phonetic realizations is allowed insofar as other linguistic information (semantic, syntactic, pragmatic, etc) makes it possible to identify a very variable sound unit. For example, the possible variability of the production of the vowel /I/ will depends on the quantity of information brougth by other linguistic fields (see figure 3).

3. A constraint-based model for domain interaction

The classical modular conception of linguistic organization considers each domain as autonomous, the interaction between them being implemented by means of relations between the different structures (for example between an intonative group and a syntactic tree). In our approach, each domain contributes partially to the construction of the information. This means that they can have a certain level of independence, but the final interpretation has to bring all pieces of information together (see [11]). As shown in the previous section, this process relies on relations that can be specified between domains.

We don't think possible, and even desirable, to propose a general and homogenous approach describing all the domains by means of a unique system. It is in our opinion preferable to describe each domain with its own system and its own formalism. But, in contrast to classical approaches, relations between domains should not be represented into a domain nor be expressed with the formalism of the concerned domains. The collaborative architecture we propose introduces a new level on top of the domains containing all interaction relations. Such relations typically consists in connecting several objects from different domains. Each object can be specified following the representation of its domain.

In the remaining of this section, we present first the notion of anchoring that makes it possible to refer to any kind of object. We describe then interaction constraints implementing relations between domains independently from their formalism. We conclude by a presentation of the general architecture of our approach that specifies the situation the different domains and their interaction levels.

3.1. Anchoring instead of integrating

Usually, relations between domains are represented into a domain (see for example [12] or [13]). We think however that the problem consists in finding an interface point between domains more than an alignment between structures. This problem has been explored in particular for multimodal corpus annotation (cf. in particular [14], [15] or [16]). In such experiments, the idea is to encode together with each piece of information its localization in the signal. In our solution, we propose a generic indexation making reference to different levels of information. For some domains (typically prosody) a temporal indexing comes naturally in mind. For some others (e.g. indexing written material), a linear indexing over the string is necessary. Finally, we also need to index information that is not usually associated with a given position but more generally with a context (see [17]). This is typically the case for discourse information. We propose then to use an anchor which is represented by a complex feature as follows:

$$\text{Anchor}\begin{bmatrix} \text{Temporal} < i,\ j > \\ \text{Position} < k,\ l > \\ \text{Context } c \end{bmatrix}$$

The temporal index is represented by two values (beginning and end). The position is also a couple of indexes (corresponding to nodes in a chart interpretation) localizing an object in the input. The *context* feature implements the notion of *universe* (i.e. a set of discourse referents) as in DRT (cf. [18]). An object can then be specified by means of different kind of information: its *domain* and its *characterization* (the set of corresponding properties) containing its *anchor*.

3.2. Interaction constraints

The description of domain interaction takes advantage of the constraint-based approach presented above. The idea is to propose a mechanism making it possible to infer new properties according to the different characterizations produced for different domains. In other words, this new kind of constraint specifies a relation between characterizations (rather than between categories). Insofar as different sources of information, coming from different domains, are involved in these relations, the characterizations have to specify the domain and the anchor. A first approximation of the interaction relation can be represented as follows:

$$\left\{ \text{Obj}i\begin{bmatrix} \text{Domain } di \\ \text{Charac } ci \\ \text{Anchor } ai \end{bmatrix},\ldots,\text{Obj}j \right\} \Rightarrow \left\{ \text{Obj}k\begin{bmatrix} \text{Domain } dk \\ \text{Charac } ck \\ \text{Anchor } ak \end{bmatrix},\ldots,\text{Objl} \right\}$$

Such a relation means that when the different characterizations Obj_i ,..., Obj_j, eventually coming from different domains, are exhibited, then the new properties stipulated in the characterizations Obj_k ,..., Obj_l are added to the general description. The general schema consists now in building characterizations of each domain and propagating new properties according to the interaction constraints. This propagation is done at the same time as the satisfaction process: new properties are propagated thanks to interaction as soon as the corresponding characterizations are instantiated. The evaluation of the interaction constraint constitutes in itself a part of a general characterization of the input. It establishes then some relations (requirement or exclusion) between categories that can have a disambiguation effect.

We illustrate in the following this aspect with an example of interaction constraints implementing the relation described in [12] and presented in the first section. It stipulates that no major breaks can separate two juxtaposed sisters connected with a complementation relation. The anchoring information allows to situate each object. This is the main interest of such a representation: an object only has to be located, its properties can be expressed independently by means of any formalism.

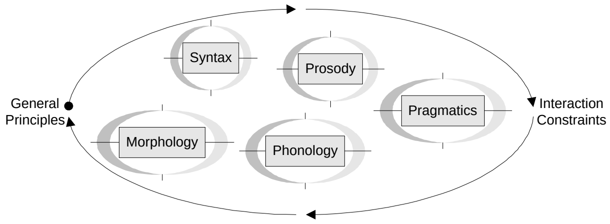
$$\begin{bmatrix} \text{Dom syntax} \\ \text{Char}\begin{bmatrix} \begin{bmatrix} \text{Cat c1} \\ \text{Anch}\begin{bmatrix} \text{Temp} < \text{t1, t2} > \\ \text{Pos} < \text{i, j} > \end{bmatrix} \end{bmatrix} \\ \begin{bmatrix} \text{Cat c1} \\ \text{Anch}\begin{bmatrix} \text{Temp} < \text{t1, t2} > \\ \text{Pos} < \text{i, j} > \end{bmatrix} \end{bmatrix} \\ \text{Comp (c2, c1)} \end{bmatrix} \end{bmatrix} \otimes \begin{bmatrix} \text{Dom prosody} \\ \text{Char}\begin{bmatrix} \begin{bmatrix} \text{Cat break} \\ \text{Anch}\begin{bmatrix} \text{Temp} < \text{t2, t3} > \\ \text{Pos} < \text{j} \ldots > \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

This interaction constraint connects two characterizations coming from the prosodic and the syntactic domains. Such interaction constraint typically works for attachment disambiguation. In case of ambiguity (for example in PP attachment), the interpretation that will be favored thanks to this constraint is the one at the higher level when a major break precedes the PP.

3.3. The general architecture

Our perspective makes it possible to consider language as a multi-layer constraint-based system. Each domain follows its own organization, possibly represented as a constraint system. We call each domain description a grammar. It is then possible to examine local constraints within a unique domain and a given formalism. The upper level of our organization contains the set of interaction constraints. It constitutes in one sense a meta-grammar over the set of domain grammars.

Each domain produces a certain amount of information. In contrast with classical approaches, being this information accessible at any time, it is also possible to apply interaction constraints at any time. This comes to produce or propagate new information which can have as a side effect a disambiguating result over the local systems.



4. Conclusion

It seems increasingly clear that the different linguistic fields interact between them in production as well as in perception of language: the comprehension of a linguistic message requires a whole of information, possibly redundant, using all the elements of the language. This process supposes that, to be understood, a message must contain a certain "quantity of information". This information is contained in the various aspects of the language (decoding acoustico-phonetics, semantics, analyze syntactic, etc). When a minimum level of information is reached by some linguistic elements, other linguistic elements are not to be informative. Thus, part of phonetic information can be imprecize if the semantic and syntactic aspects are sufficiently informative. This project comes within a general theoretical step aiming at clarifying the functional links which interact between the linguistic fields. We think that this approach can enable us to understand how and why, on each level of analysis, linguistic information can be unstable, partial even absent.

5. References

[1] Croft W. & D. Cruse (2003) *Cognitive Linguistics*, ms, Cambridge University Press.

[2] Langacker R. (1999), *Grammar and Conceptualization*, Walter de Gruyter.

[3] Meunier C., Floccia, C. (1997) "La spécificité des types de parole pour la perception de la voyelle", Actes des Journées d'Etudes Linguistiques "La voyelle dans tous ses états", Nantes, 158-163.

[4] Meunier C., Floccia, C. (1999) "Syllabe ou mot : quelle unité permet d'identifier les catégories phonétiques?", *Actes des Journées d'Etudes Linguistiques "Syllabes"*, Nantes, pp. 87-92.

[5] Stevens K.N. and Blumstein S.E. (1978), Invariant cues for place of articulation in stop consonants, J.A.S.A., 64, pp. 1358-1368.

[6] Blumstein. S.E (1986) "On acoustic Invariance in Speech" in Invariance and variability in speech processes, Perkell and Klatt eds., M.I.T., London.

[7] Shannon, C.E., and Weaver, W. (1949) The mathematical theory of communication, Urbana, IL: University of Illinois Press.

[8] Lindblöm, B. (1990), "Explaining phonetic variation: a sketch of the hyper- and hypospeech theory", *Speech Production and Speech Modelling*, Hardcastle and Marchal (eds.), Kluwer Academic Publishers, 403-439.

[9] Ganong, W.F. (1980), “Phonetic categorization in auditory word perception”, *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110-125.

[10] Warren , R.M. (1970) "Perceptual restoration of missing speech sounds", *Science*, 167, 392-395.

[11] Kettebekov S. M. Yeasin & R. Sharma (2002),"Prosody Based Co-analysis for Continuous Recognition of Coverbal Gestures", in *proceedings of International Conference on Multimodal Interfaces*.

[12] Bear J. & P. Price (1990), “Prosody, Syntax and Parsing'”, in proceedings of the *ACL-90*.

[13] Hirst D. (1993), "Detaching Intonational Phrases from Syntactic Structure", in *Linguistic Inquiry*, 24:4.

[14] Bird S. & al. (2000) ``ATLAS: A flexible and extensible architecture for linguistic annotation", in *proceedings of LREC-2000*.

[15] Bird S. & M. Liberman (2001) "A formal framework for linguistic annotation", in *Speech Communication*, 33:1.

[16] Blache P. & D. Hirst (2001) "Aligning Prosody and Syntax in Property Grammars.", in *proceedings of Eurospeech-2001*.

[17] Pineda L. & G. Garza (2000), "A Model for Multimodal Reference Resolution", in *Computational Linguistics* 26:2.

[18] Kamp H. & U. Reyle (1993) *From Discourse to Logic*, Kluwer.