

# The Weight of Phonetic Substance in the Structure of Sound Inventories

Nathalie Vallée, Louis-Jean Boë, Jean-Luc Schwartz, Pierre Badin, Christian Abry  
Institut de la Communication Parlée, UMR CNRS 5009, INPG – Université Stendhal  
BP 25 X – 38040 Grenoble Cédex 9, France

## Abstract

In the research field initiated by Lindblom & Liljencrants in 1972, we illustrate the possibility of giving substance to phonology, predicting the structure of phonological systems with *non-phonological* principles, be they listener-oriented (perceptual contrast and stability) or speaker-oriented (articulatory contrast and economy). We proposed for vowel systems the Dispersion-Focalisation Theory (Schwartz et al., 1997b). With the DFT, we can predict vowel systems using two competing perceptual constraints weighted with two parameters, respectively  $\lambda$  and  $\alpha$ . The first one aims at increasing auditory distances between vowel spectra (dispersion), the second one aims at increasing the perceptual salience of each spectrum through formant proximities (focalisation). We also introduced new variants based on research in physics - namely, phase space  $(\lambda, \alpha)$  and polymorphism of a given phase, or superstructures in phonological organisations (Vallée et al., 1999) which allow us to generate 85.6% of 342 UPSID systems from 3- to 7-vowel qualities. No similar theory for consonants seems to exist yet. Therefore we present in detail a typology of consonants, and then suggest ways to explain plosive vs. fricative and voiceless vs. voiced consonants predominances by i) comparing them with language acquisition data at the babbling stage and looking at the capacity to acquire relatively different linguistic systems in relation with the main degrees of freedom of the articulators; ii) showing that the places “preferred” for each manner are at least partly conditioned by the morphological constraints that facilitate or complicate, make possible or impossible the needed articulatory gestures, e.g. the complexity of the articulatory control for voicing and the aerodynamics of fricatives. A rather strict coordination between the glottis and the oral constriction is needed to produce acceptable voiced fricatives (Mawass et al., 2000). We determine that the region where the combinations of Ag (glottal area) and Ac (constriction area) values results in a balance between the voice and noise components is indeed very narrow. We thus demonstrate that some of the main tendencies in the phonological vowel and consonant structures of the world’s languages can be explained partly by sensorimotor constraints, and argue that actually phonology can take part in a theory of Perception-for-Action-Control.

## 1 Phonology in a substance-based linguistics

Speech communication operates on two highly-structured levels, the system itself and its physical realisation. This is probably the reason why speech communication is so efficient compared to other communication means used by man or animal. The terms *language* and *speech* refer to these two levels, separated by Saussurean structural linguistics in *form* and *substance*, and reconsidered by generative grammar under the terms *competence* and *performance*. Throughout the 20th century, several axioms of the core of structuralist, and subsequently generativist, approaches have conditioned relationship between phonetics and linguistics:

- the language/speech dichotomy;
- the independence of these two concepts;
- the primacy of language over speech.

This distinction is the result of a particular methodological approach. Linguistics, in order to make empirical data intelligible, separate the study of a sound system – its field of research – from the issue of its physical realisation, which may be variable and polymorphous (Ducrot & Schaeffer, 1995, p. 245). These methodological principles, cumulative in their effects, marginalised any attempt to reveal interactions between the major tendencies observed in the phonological systems of the world's languages – their universals – and the articulatory and acoustic characteristics of their physical realisation. They isolated linguistics in a reductionist internalism and influenced the presuppositions which founded phonology. According to these principles, phonological units cannot be defined by substantial properties but only with respect to their relative position within the system, and the question of their universality no longer arises. On the threshold of the 21st century, a number of approaches in contemporary linguistics and phonology are still characterised by a strong internalist approach, often presented as an advantage, and goes as far as outright refusal to take into account hypotheses, models, and results obtained by connected disciplines which have language and speech in their field of research. This rejection – to consider the evidence of relationship between form and substance – reiterated throughout this century, is perhaps a unique example in the history of the twentieth-century science: testing data and models, whatever their provenance, should form an intrinsic part of any scientific approach.

In 1952, Jakobson, Fant, and Halle introduced, in *Preliminaries to Speech Analysis* (PSA), a new conception of phonology in linking phonemic features to acoustic correlates and speech perception. Even if their proposed features – which were too general and poorly quantified – have not really clarified the relation between form and substance, the relationship between phonology and phonetics interrupted for almost two decades was discussed anew.

Generative phonology retained from PSA the idea of a universal system of binary features. In *The Sound Pattern of English* (SPE) in 1968, Chomsky and Halle replaced the traditional acoustico-perceptual specification by a universal phonetic representation, expressed in terms of more numerous articulatory features which were precisely defined and well-documented. This can be considered as a very important advance in the framework of phonetic description and relationship between form and substance. It might have been expected that generativist phonologists could have connected their work to articulatory measurements. In fact, the descriptive prolegomena of SPE have remained unfollowed, and the proposed features have not been used in phonological descriptions as a part of a pure symbolic formalism, as the authors themselves stated (Chomsky & Halle, 1968, p. 274). To compensate the lack of naturalism, Chomsky and Halle reintroduced the Theory of Markedness (inherited from the phonology of Trubetzkoy). More recently, the Optimality Theory was proposed to preserve those universal constraints which reveal the unity of language. For Prince and Smolensky (1993), the universal grammar can be essentially considered as a set of ordered constraints, often conflicting, which regulate the well-formedness of representations from which individual grammars are constructed. These constraints are always active, and languages are then distinguished by the way in which conflicts are resolved. Once again, however, it is necessary to proceed with criteria which do not reduce the reasoning to a straightforward tautology.

In fact, the Saussurean dogma taken up by Chomsky (1965): “The classical Saussurean hypothesis of the logical priority of the study of language (and of the generative grammar that describes it) seems almost incontrovertible” has not truly been called into question, except in 1972, with a new perspective that paved the way towards a whole new sweep of research called “substance-oriented” linguistics brought out by The Maximal Dispersion Theory (Liljencrants & Lindblom, 1972) and The Quantal Theory (Stevens, 1972). We place ourselves in these two approaches initiated in the seventies, using recent work on

the typology of sound structures (Vallée, 1994 ; Schwartz & al. 1997a ; Stefanuto & Vallée, 1999), and trying to show that some of the tendencies refer to biological constraints on speech production and speech perception human systems, that is, to the substance, not to the form. Our aim is not to refute the existence of a formal phonological level with its intrinsic formal principles and rules, but to try to determine, and if possible quantitatively model, a set of constraints coming from the speech substance and capable of having played a part in the emergence of this formal system – and therefore to throw some lights on phonological facts which could sometimes appear arbitrarily.

Since Lindblom, a number of elements are now available to integrate phonology in a substance-based theory called the Perception-for-Action-Control Theory (Schwartz et al., 2002): *“this theory should be able to show how the choice of speech units inside the phonological system may be constrained and patterned by the inherent limitations and intrinsic properties of the speech perception system – and its indissociable companion, the speech production system”*. The core of the proposal is that *“a listener might follow the vocalisations of his speaking partner, in order perhaps to understand them, but at least certainly to imitate and learn: in other words, perception enables a listener to specify the control of his future actions as a speaker. On the other hand, [...] the perceptual representations of speech gestures transform, deform, shape the speaker's gestures in the listener's mind, and hence provide templates that in return also help to specify the control of the speaking partner's own actions.”*

This approach is centred on the co-structuring of the perception and action systems in relation with phonology. However, the Perception-for-Action-Control Theory does not fall within the framework of both *“an "auditory" theory in which the sensory-interpretative chain is considered independently of the patterning of sounds by speech gestures, in the search of some "direct link" between sounds and phonemes; and from a "motor" theory [...] in which perception is nothing but a mirror of action, in the claim of a direct link between sounds and gestures.”*

The studies presented here are at the core of the relationship between phonology and perception for action control. We attempt to show that phonemes, vowels and consonants, are not obviously arbitrary phonological units: phonological systems are in part co-structured by speech perception and action. Considering that phonology should contain the set of formal structures characterising conscious mechanisms for speech control, it is only logical to assume that it is not independent of the ability of the speech production system to produce gestures, and of the speech perception system to recover and shape these gestures. This is the basis both of a theory we have developed for dealing with oral vowel systems, in the line of Lindblom's Dispersion Theory: The Dispersion-Focalisation Theory (DFT) presented in the following section, and of a set of suggested ways to explain plosive vs. fricative and voiceless vs. voiced consonants predominances then developed from our typological analysis based on UPSID phonological systems. In fact, we adopt an epistemological framework using “external” data to phonological description: speech production and speech perception constraints to which it is possible to add some data on ontogenesis and language disturbances (cf. MacNeilage, 1998). Following this approach initiated by Lindblom, models currently permit the prediction of the main tendencies observed in sound systems. It is thus possible to take a close look at the problem of phonological structures and their changes systematically, to establish a precise diagnosis of what can be attributed to speech production/perception, and to list the questions which must be addressed to linguistics and sociolinguistics instead. With such an approach, we do not fall into the trap of the weaknesses of an inductive approach which consists in inferring general laws from isolated observations and can lead us to the error of presupposing the conclusion. We finally illustrate and discuss the inescapable fact

that the relationship between phonology and phonetics has to constitute a research field of linguistic sciences.

## 2 The weight of phonetic substance in vowel structures

### 2.1 Prediction of the phonological structure of vowel systems: the DFT

#### 2.1.1 General principles

Since the beginning of the 70s, several proposals have been made to predict the phonological structure of vowel systems with *non-phonological* principles, be they listener-oriented (perceptual contrast and stability) or speaker-oriented (articulatory contrast and economy). The so-called “sufficient perceptual contrast” theory (Lindblom, 1986) provides the best global fit with phonological data. However, to overcome its two main problems (that is, the excessive number of high non-peripheral vowels in the model predictions and the impossibility to predict the [i y u] series within the high vowel set), we proposed at ICP a theory based on two principles, that is dispersion and focalisation. These principles specify two basic properties that vowel gestures should have in order to provide a viable sound system for communication. Firstly, gestures should provide sufficiently different acoustic patterns to allow the perception system to be able to recover them without confusions or ambiguities: this is dispersion. Secondly, they should provide salient spectral patterns (formant convergence in vowel spectra), easy to process and characterise in the ear: that is focalisation. While auditory dispersion is a classical concept, focalisation is a principle introduced by ourselves (Schwartz & Escudier, 1987, 1989). The Dispersion-Focalisation Theory (DFT) (Schwartz et al., 1997b) allows us to predict vowel systems through a competition between two perceptual costs: for a given number of vowels, the most frequent system in the world’s languages is supposed to be obtained by minimising a global criterion combining a structural dispersion cost and a local focalisation cost.

#### 2.1.2 Implementation

Each vowel is characterised by the formants of its spectrum, that is  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$ , expressed in a perceptual Bark scale. The ( $F_2, F_3, F_4$ ) set allows to compute an integrated “effective perceptual formant”  $F'_2$ . In the Dispersion-Focalisation Theory (DFT), we define a vowel system by a set of vowels in the maximum available formant space and we associate to each system an energy function consisting of the sum of two costs, namely a structural dispersion cost based on inter-vowel perceptual distances – computed through an Euclidean distance in the ( $F_1, F'_2$ ) space, and favouring large inter-vowel distances – and a local focalisation cost based on intra-vowel perceptual salience, which aims at providing perceptual preference to vowels showing a *convergence between two formants*, that is, vowels with close  $F_1$  and  $F_2$ ,  $F_2$  and  $F_3$ , or  $F_3$  and  $F_4$ . The model is controlled by two parameters:  $\lambda$  specifying the weight of  $F'_2$  in respect to  $F_1$  in the dispersion cost, and  $\alpha$  specifying the respective weight of the focalisation cost relative to the dispersion cost. Then, for a fixed number of vowels in a system, we implemented various algorithms to select optimal systems, that is systems with the lowest energy (the best compromise of dispersion and focalisation), either locally (“stable systems”) or globally (“best systems”) (Schwartz et al., 1997b). Our predictions of optimal vowel systems were then systematically compared to vowel inventories, according to the UCLA UPSID Database (Maddieson, 1984 ; Maddieson & Precoda 1989).

### 2.1.3 Phase spaces

For a given number of vowels, from 3 to 9 (beyond this limit, vowel systems introduce a new dimension, mainly nasality and less often quantity, Vallée, 1994), we can predict, in the DFT framework, different vowel systems in the  $(\lambda, \alpha)$  space. This leads to the determination of what we call “the phase space”, a well-known procedure in thermodynamics used to predict the states of a substance (such as the states of water: steam, liquid and ice), as a function of pressure and temperature. The general trend is that, for a given number of vowels in a system, decreasing  $\lambda$  favours peripheral systems while increasing it favours systems with one and then two high non-peripheral vowels; and increasing  $\alpha$  favours focal vowels, and particularly stabilises [y] within an [i y u] high series, while this series is unstable when  $\alpha$  is set to 0.

Previous work allowed us to verify that these predictions were more or less compatible with the observed preferred phonological vowel systems in the UPSID<sub>317</sub> database (Maddieson, 1984).

Considering that peripheral systems are generally preferred from 3 to 7 vowels and that the [i y u] series of high vowels exists in a significant amount of cases in the database (about 5% of the cases in the whole database, and 13% of the cases for systems with 7 vowels or more), we showed that setting the  $\lambda$  value around 0.2 - 0.3 and the  $\alpha$  value around 0.3 - 0.4 led to quite acceptable predictions (Schwartz et al., 1997b). In the present work, we try to go one step further: we shall attempt to determine where in the phase spaces one can find the different systems, *preferred or not*, existing in UPSID<sub>451</sub>, and what kinds of “superstructures” can be derived from this analysis.

### 2.1.4 Structural symmetries between vowel systems: a typological equivalence criterion

#### 2.1.4.1 Prototypical structures in phase spaces

Our previous simulations led to “prototypical systems”. These are winning n-vowels systems in the DFT framework, in the sense that they have a minimal global Dispersion-Focalisation (DF) energy, according to the values of the two free parameters  $\lambda$  and  $\alpha$ .

We have focused our study on values of n from 3 to 7 because they allow us to capture the most significant phonological tendencies of the UPSID database. The DFT simulation results are given in Figures 2-6, respectively for n = 3, 4, 5, 6 and 7. For each value of n, the phase space determines regions in the  $(\lambda, \alpha)$  space in which a given system wins (with its vowel qualities displayed as black points on a prototypical grid). We see that there are two prototypical systems for n = 3, which we call  $S_3T_1$  and  $S_3T_2$ . There are four prototypical systems for n = 4, 5, 6, and five prototypical systems for n = 7; let us call them  $S_nT_i$ , with n from 3 to 7, and i from 1 to 5. The global trend is that increasing n increases the dispersion cost of peripheral systems, hence it decreases the  $\lambda$  boundary necessary for making these systems optimal. Hence peripheral systems are favoured with small values of  $\lambda$ . When  $\lambda$  is too small, the vowel space is completely vertically stretched (since higher formants play a minimal part in the determination of vowel phonetic quality); this favours asymmetrical peripheral configurations because of the interactions between front and back peripheral vowels in the systems. Non-peripheral configurations, that is systems with more than two high vowels, appear with large  $\lambda$  values, and when  $\alpha$  increases, focal vowels (especially [i] with close  $F_3$  and  $F_4$ , other front unrounded vowels together with [y], all with close  $F_2$  and  $F_3$ , and back rounded vowels, with close  $F_1$  and  $F_2$ ) are favoured. Decreasing  $\alpha$  leads to replacement of the high rounded vowel [y] with a high vowel acoustically more central (namely [i] or [u]).

### 2.1.4.2 Reverse prototypical structures

We hypothesised that two structures having the same number of peripheral vowels but systematically replacing front unrounded vowels by back rounded ones with the same height, and vice-versa, are equivalent structures in the sense of DFT, that is to say that they have roughly the same DF energy for a given value of  $n$  and of the  $(\lambda, \alpha)$  pair. This was systematically verified by comparing the energy of the  $S_n T_i$  prototypical systems with reverse systems that we called  $S_n T_i^*$ . For example, for  $n=4$  we compared  $S_4 T_1 = [i o \varepsilon a]$  with  $S_4 T_1^* = [i e \circ a]$ , and  $S_4 T_2 = [i 'e' a u]$  with  $S_4 T_2^* = [i 'o' a u]$ ,  $S_4 T_3 = [i y u a]$  and  $S_4 T_4 = [i i u a]$  having no reverse counterpart. Indeed, we confirmed that  $S_n T_i^*$  structures have a DF energy quite close to the  $S_n T_i$  ones whatever the region of the phase space, that is to say whatever the  $\lambda$  and  $\alpha$  values. Pushing the analogy with physics one step further, this reminds us of the “polymorphism” of a number of solids (e.g. metals, or crystals). In this situation, while fusion produces a homogeneous liquid phase, solidification leads to mixtures of two or more variants of the solid phase, all stable and more or less with the same energy. This is exactly what happens here with the two variants within a given phase. Hence, our typologies of phase spaces involve “superstructures” grouping prototypical structures and reverse ones (displayed with white points instead of black ones in Figures 2-6). The relevance of these superstructures for describing the UPSID database will now be discussed in the next section.

### 2.1.5 Comparing UPSID data with DFT simulations

#### 2.1.5.1 UPSID data reanalysed

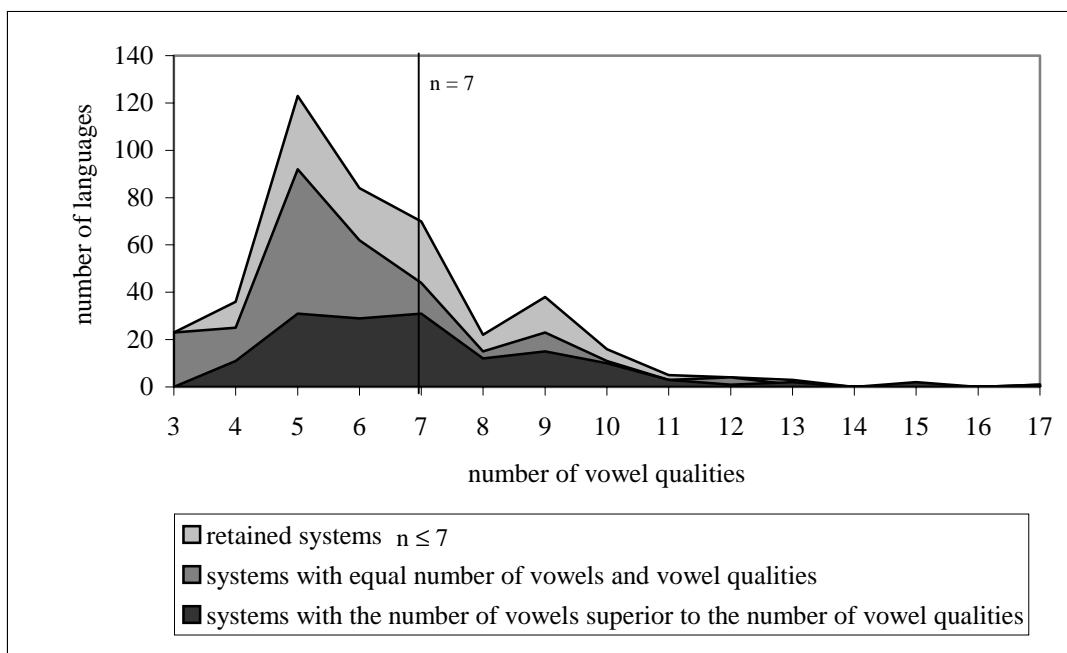
Since Trubetzkoy and his *Principles* (1939), taxonomy has not only been an approach of historical linguistics: associated to research on synchronic trends, it constitutes today a main stage in linguistic theories. Institutionalised in 1961, under the aegis of the *Social Science Research Council* during the *New-York Conference on Language Universals*, this research field aims at finding common basic structures in languages – in diachrony as well as in synchrony. *The Language Universals Project* (1967-1976) led to the building up of the *Stanford Phonology Archives* (Greenberg et al., 1978), with which many important studies dealing with typological classification and phonological tendencies, were achieved (Sedlak, 1969; Crothers, 1978; Maddieson, 1984; Vallée, 1994). But all these studies present variegated contents: data are constantly enriched, questions on the materials vary from one author to the other. The UPSID (UCLA Phonological Segment Inventory Database) (317 then 451 languages) gathers phonological systems of languages in the world, sampling more or less uniformly all linguistic families. UPSID<sub>317</sub> (Maddieson, 1984), then UPSID<sub>451</sub> (Maddieson & Precoda, 1989) were chosen to approximate a properly constructed quota sample on a genetic database of the world’s existing languages.

UPSID was implemented at ICP several years ago and we have been using it for vowel and consonant research. In order to test our hypotheses, we have reanalysed the UPSID database of vowel systems, thanks to a two-step methodology.

The languages in UPSID have 3 to 28 vowels. Firstly, from raw data, that is to say without any typological equivalencies, we obtain 252 types of phonological structures from 3 to 17 vowel qualities. What we call vowel qualities corresponds to “basic segments” (vs. “elaborated” and “complex” segments) in the sense of Lindblom & Maddieson (1988). We note that more than 96% of the languages have from 3 to 10 basic vowel qualities, and if we focus our study on systems with 3 to 7 qualities, we obtain 77% of the 451 languages (348 systems). This is due to the fact that there are in many cases more vowels than vowel qualities

in a given system; for instance /i e a o u/ is the phonological structure of four UPSID languages of which three have more than 5 vowels: Chipewyan with 14 vowels /i e a o u i: a: u: ɨ̃ ɛ̃ ɔ̃ ɨ̃: ɛ̃: ɔ̃:/, Siriono (12) /i ɨ̃ e ɛ̃ a ɛ̃ o ɔ̃ u ɨ̃ iʸ ɨ̃ʸ/, Tamang (10) /i e a o u ɨ̃ ɛ̃ ɔ̃ ɔ̃ ɔ̃ ɔ̃/. The systems with nasal, laryngeal, pharyngeal or retroflexed vowels sharing no vowel qualities with a basic segment, as opposed to the systems quoted above, have been discarded for follow-up analyses. These results in eliminating less than 3% of UPSID's languages and 3.4% of languages having from 3 to 7 vowels, that is seven languages with 6 vowel qualities and five languages with 7 qualities, for instance the Cherokee system /i 'e' 'ə' a 'o' u/ or the Tarascan system /i ɨ̃ e a o u/. At this stage we retain 336 systems of the database.

Secondly, we take into account the so-called “transparency rule” (Schwartz et al., 1997a). This rule states that schwa should be conceived as a separate class, considering that it does not seem to interfere with the other vowels in a system: indeed, schwa added or removed from a system does not disrupt the structural organisation of this system. The “transparency rule” concerns 64 languages from 4 to 8 vowel qualities. For instance we have classed the Ivatan structure /i 'ə' a u/ as  $S_3T_2$ , Achumawi /i 'e' 'ə' a 'o' u/ as  $S_5T_2$ , Ndut /i e ɛ̃ a ɔ̃ o u/ as  $S_6T_1$ , and Fur /i e ɛ̃ 'ə' a ɔ̃ o u/ as  $S_7T_2$ . The “transparency rule” results in slightly increasing the number of systems in the analysis, thanks to six 8-vowels systems which become 7-vowels ones. Hence at this last stage we stay with 342 systems that is to say 75.8% of the database (Figure 1).



**Figure 1:** UPSID's languages distribution by number of basic vowel qualities. We focus our study on systems from 3 to 7 qualities.

### 2.1.5.2 Distribution of UPSID data within phase spaces

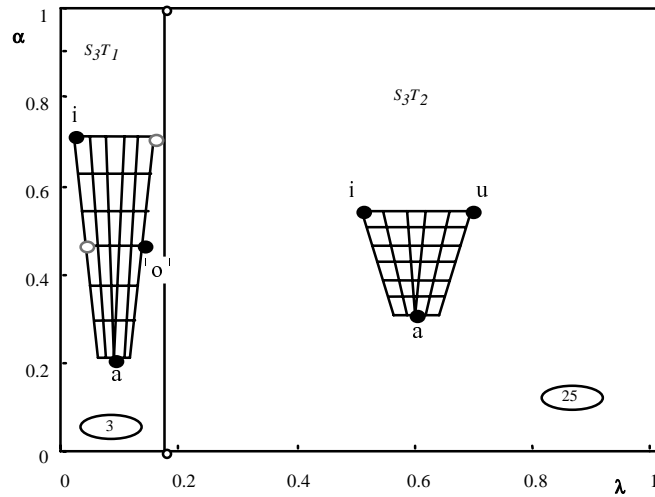
We now have at our disposal both a series of predictions organised around the typology  $S_nT_i / S_nT_i^*$  defined in Section 2.1.4, and an inventory of 342 systems (the three-fourths of UPSID<sub>451</sub>) with 3 to 7 vowel qualities. The final goal of this work was to try to associate to most of these 342 systems a region in the phase space where they would be optimal (i.e. viable in the sense of the DFT). This is displayed in Figures 2-6, where we have plotted within each region of the phase spaces the number of systems fitting with the corresponding structure. Let us now discuss the obtained results in more detail.

To begin with, it appears that 303 of the 342 3-to-7 vowel systems (88.6%, or 67.2% of the whole UPSID<sub>451</sub> database) fit with one of the  $S_nT_i$  or  $S_nT_i^*$  types. The 39 rejected systems (fitting with no prototypical or reverse type) correspond to one system in  $S_3$ , 5 in  $S_4$ , 9 in  $S_5$ , 10 in  $S_6$ , 14 in  $S_7$ ; hence their number increases with  $n$ , which is logical since the complexity of the distribution of vowel qualities increases.

Next, the most widespread types in Figures 2-6 are those corresponding to  $S_nT_2$  (and sometimes to  $S_nT_2^*$ ). This provides a first confirmation on UPSID<sub>451</sub> of our results on UPSID<sub>317</sub> (Vallée, 1994), namely that the  $(\lambda, \alpha)$  region defined by  $0.2 \leq \lambda \leq 0.3$  and  $0 \leq \alpha \leq 0.4$  is compatible with preferred 3-to-7 vowels systems in UPSID. But the data in Figures 2-6 provide some new confirmation of this result. Indeed, it appears that systems corresponding to types associated to large  $\lambda$  or  $\alpha$  values are quite few. On the contrary, most systems on Figures 2-6 are located at low  $\lambda$  values. Indeed, apart from the “best”  $S_nT_2$  structures, other structures generally occupy nearby regions (mostly of types  $S_nT_1$  or  $S_nT_1^*$ ) and if we define a broad acceptable region such as  $0.1 \leq \lambda \leq 0.3$  and  $0 \leq \alpha \leq 0.4$ , we obtain a total of 293 systems, that is 85.6% of the 342 systems of our inventory, which is quite important. Altogether, this confirms with a strong reliability the need to “stretch” the acoustic space along the  $F_1$  dimension in auditory spectral distances, which indicates the dominant role played by the lower formant  $F_1$  in vowels’ phonetic quality.

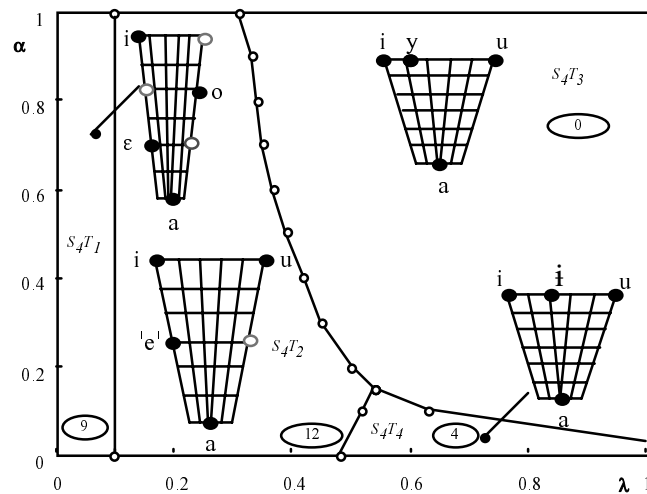
The next observation deals with the symmetry between front and back peripheral vowels. Globally, the data in Figures 2-6 confirm the well-known fact that vowel systems “prefer” both peripheral vowels and front-back symmetry. In the asymmetrical cases, when the numbers of front and back vowels are different, the (classical) trend is that there are more front than back ones: for example, 9  $S_4T_1$  [i o ε a] vs. 0  $S_4T_1^*$  [u e ɔ a] structures, 8  $S_4T_2$  [i u 'e' a] vs. 4  $S_4T_2^*$  [i u 'o' a] structures, 14  $S_6T_1$  [i u e 'o' ε a] vs. 10  $S_6T_1^*$  [i u o 'e' ɔ a] structures. When the number of front and back vowels are the same, the (less classical) trend is that front vowels have often a more open degree than back ones; though this is not true for 3-vowels systems (2  $S_3T_1$  [i 'o' a] vs. 0  $S_3T_1^*$  [u 'e' a] structures), it is clearly the case for 5-vowels systems (4  $S_5T_1$  vs. 16  $S_5T_1^*$ ) and for 7-vowels systems (0  $S_7T_1$  vs. 2  $S_7T_1^*$ ).



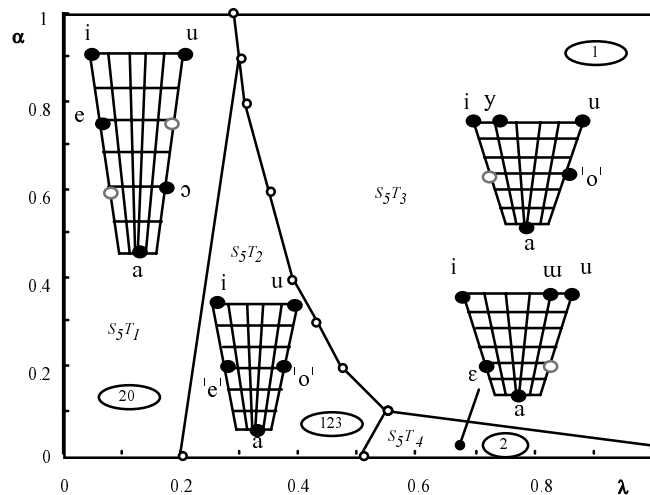


**Figure 2:** Phase space for three-vowels systems.

$S_n T_i$  structures have their vowel qualities displayed as black dots on a prototypical grid.  $S_n T_i^*$  vowels are displayed as white dots replacing the black ones. The number in the oval is the total number of UPSID languages with the  $S_n T_i$  and  $S_n T_i^*$ .



**Figure 3:** Phase space for the four-vowels systems.



**Figure 4:** Phase space for the five-vowels systems.

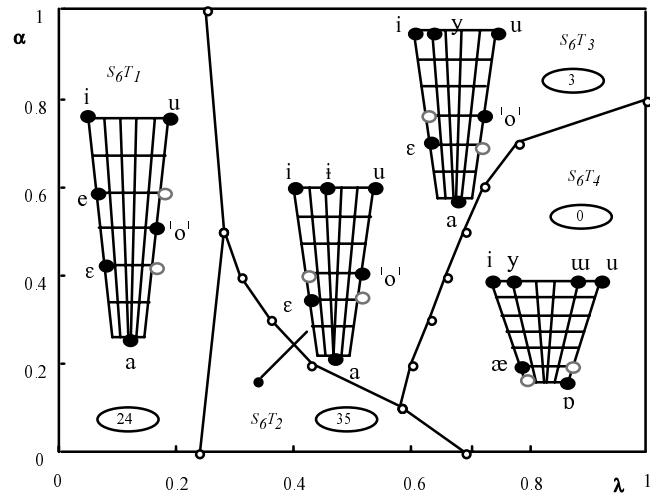


Figure 5: Phase space for the six-vowels systems.

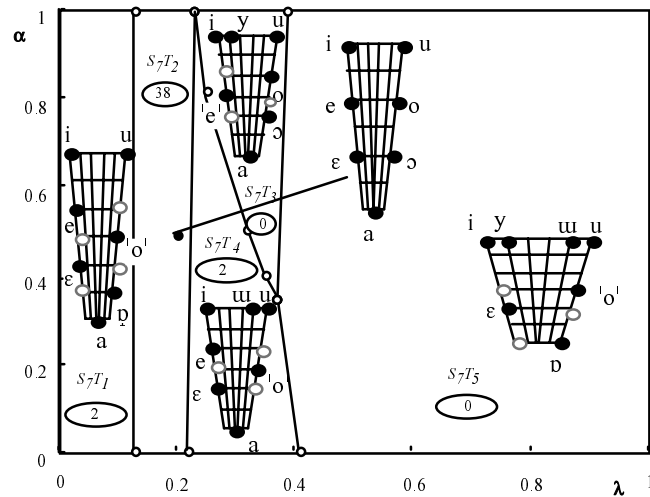


Figure 6: Phase space for the seven-vowels systems.

In what concerns focalisation, its role is more important for stabilising the vowel structures of larger systems, particularly those containing [y]: this vowel appears in only 2% of UPSID's languages from 3 to 7 vowel qualities, but almost 7% of all languages; notice that more than half of them are Indo-European and Uralo-Altaic languages (Alcantara, 1998). In our simulations, /y/ is only present in the  $S_6T_3$  [i y u 'o' ε a] structure (3 examples, that is only 1% of our reduced database).

### 3 Consonant systems and some substantial constraints on form

Whereas we have clearly established a theory for the prediction of vowel systems, no similar theory for consonants exists at present. Therefore we will first describe in detail a typology of consonants, and then suggest ways to explain some of the observed tendencies.

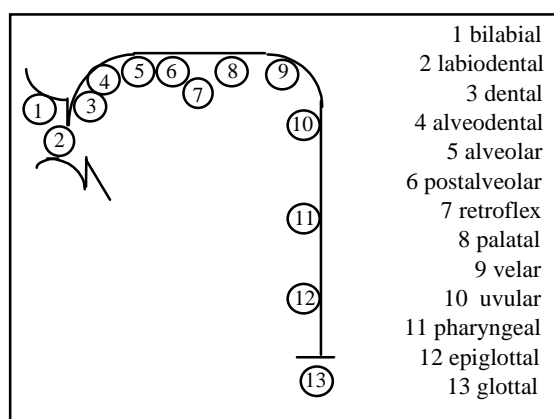
### 3.1 Typology

#### 3.1.1 Taxonomic elements

Since the 1960's, typological works have dealt more with vowels than with consonants. This is partly due to the wide range of consonant inventory sizes and the large number of classification parameters that make type emergence difficult. Although the nature of some “external” constraints influencing the content of vowel systems is coming to light (Schwartz et al., 1997a), we are still far from understanding the biological constraints influencing the consonant structures of the world's languages.

Investigating existing languages is the traditional way to search for universal tendencies, but we also need to look at the question of the capacity to acquire relatively different linguistic systems. Firstly we present a set of consonant inventory tendencies, then we compare them to available language acquisition data – more precisely, at the babbling stage. Indeed, among the different ontogenetic stages, babbling seems to be essential since it marks “precursors” of universal speech attributes, following MacNeilage's Frame then Content Theory (MacNeilage & Davis, 1990; MacNeilage, 1998).

Basically, our typological study is built on places and manners of articulation, two very usual parameters in consonant classification. UPSID<sub>451</sub> lists 920 phonemes with 654 consonant segments distributed across 13 places (Figure 7) and 16 manners.



**Figure 7:** Consonant articulation places (UPSID<sub>451</sub>).

In our typology, manners have been grouped into 7 types: plosives (with implosives, ejectives, glottal stops); nasals; fricatives (with ejectives and *h-sounds*); affricates (with ejectives); approximants; trills/taps/flaps; clicks.

Compared to UPSID, our place classification, based on Creissels (1994) and Ladefoged & Maddieson (1996) is more detailed (17 simple articulation places and 5 double articulation places).

We have kept the retroflex category as a place, as suggested in Maddieson (1984), Ladefoged & Maddieson (1996) and the IPA (1996), though it refers to a manner: the distribution of retroflex consonants across their different places of articulation (alveodental, alveolar, prepalatal, etc.) would require an investigation of the sources of transcription.

We have to stress that making the comparison to previous taxonomies (let us mention works by Hockett (1955), Hagège (1982), Maddieson (1984), Lindblom & Maddieson (1988), Laver (1994)) is not easy in so far as they all differ in their data preparation and classification methodologies.

### 3.1.2 Distribution

UPSID<sub>451</sub> languages most often use from 18 to 25 consonants: minimum 6 for Rokotas (a Papuan language, 11 phonemes), maximum 95 (with 48 clicks) for !Xũ (Khoisan family, 141 phonemes). The average size of consonant inventories is 22, with 7.8 plosives, 4.1 fricatives, 3.3 nasals, 2.9 approximants, 2 affricates, 0.6 ejective, 0.5 trill/flap, and 0.2 click.

Table I shows the distribution of consonants across the different places of articulation. Grouping all manners, alveodentals are the most numerous (15.3%), followed by bilabials (14.3%), and velars (12.6%). It is important to note that the coronal class (*i.e.* dental, alveodental, alveolar, postalveolar and retroflex consonants grouped together) represents far and away the largest class of segments, with 44.5% in UPSID<sub>317</sub>, far ahead of bilabials. Though the coronal class represents a wide range of articulation types (Ladefoged & Maddieson, 1996), its use in our typology is acceptable: in inventories, coronal place features seldom serve as phonological distinctions between consonants of this type, except for affricates, and never for anterior coronals (Keating, 1990).

The UPSID<sub>317</sub> consonant distribution shows that some manners are more represented than others: plosives 38.6%, fricatives 20.2%, nasals 14.6%, approximants 13%, affricates 9.6%, trills/taps/flaps 3.9%. Voiceless oral coronal plosives (type /t/) are the most frequent (they are present in 97.5% of languages); bilabial and coronal nasals (types /m/ and /n/) exist in more than 9 languages out of 10; /k/, /j/, /p/ exist in more than 80% of languages; /w/ and /s/ in 2 out of 3 languages; /d/, /b/, /h/, in more than 60% of languages; one language in 2 has /l/, /g/, /ŋ/, /ʀ/.

alveodental	15.3%
bilabial	14.3%
velar	12.6%
palatal	10%
apico-alveolar	9.9%
glottal	7.6%
lamino-postalveolar	6.7%
labiovelar	5.1%
dental	5%
lamino-alveolar	3.4%
labiodental	3.3%
retroflex	2.9%
uvular	1.7%
<i>others</i>	<1%

**Table I:** Frequency of occurrences for consonant articulation places from UPSID<sub>317</sub> (Stefanuto, 1996).

### 3.1.3 Systems

#### 3.1.3.1 Size of articulation place systems per articulation manners

If the study of the size of consonant inventories gives very little information about their content, contrary to vowels (Schwartz et al., 1997a), the development of a place/manner typology reveals a strong correlation between (i) the size of place systems and (ii) the different articulation manners (Table II):

- In the majority of languages, plosives and nasals occupy, respectively, from 3 to 6 and from 1 to 6 places. The plosive manner is the only one that sound systems universally exploit. Moreover, it is also the only one that systematically uses place contrasts: languages distinguish at least 3 of them. Compared with fricatives, stops (including nasals) are almost always more numerous (see Figure 8).
- Fricatives and approximants are the consonant classes which may recruit the greatest number of place contrasts: fricatives are more widely distributed on the different system sizes. Fricatives have systems of contrasts often more complex than plosives. 8% of languages do not involve fricatives at all (contrarily to plosives, which are universally used).
- For trills, flaps and taps, systems hardly ever present a place contrast – and never spread over more than 2 places. We find the same trends in affricates.

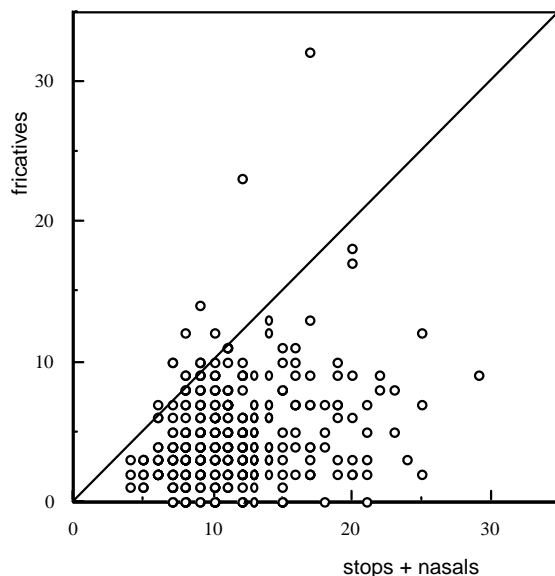
More broadly speaking:

- Languages systematically use place contrasts for plosive consonants; almost one language in two distributes its plosives in a 4-places system.
- Place contrasts are rare for affricates, and exceptional for trills/taps/flaps.
- Approximants frequently spread among 3 articulation places.
- Fricatives spread more widely across the different sizes of place systems.

Places number	1	2	3	4	5	6	7	8
trills, flaps, taps %	<b>67.2</b>	4.7						
Affricates %	<b>44.8</b>	20.8	6.9	0.3				
Plosives %			28.4	<b>43.2</b>	22.4	5.9		
Nasals %	2.5	<b>31.5</b>	30.6	25.9	4.5	2.2		
Approximants %	8.2	26.5	<b>49.8</b>	8.2	2.5	0.3	0.3	
Fricatives %	3.4	18.3	<b>25.9</b>	21.1	14.2	7.3	1.9	0.9
Place opposition Number	-	1	3	6	10	15	21	28

**Table II:** Size of the place systems in UPSID<sub>317</sub>.

The number of place oppositions changes according to manner; from 0 (*i.e.* 1 place per system) to 28 (8 places). The presented values do not include the voicing feature (voiced and voiceless categories are grouped). 43.2% of the 317 languages have a 4-places plosive system, and 49.8% have approximants spreading on 3 places.



**Figure 8:** Quantity of fricatives regarding quantity of stops plus nasals in UPSID<sup>451</sup>'s phonological systems.

The line corresponds to languages with an equal number of stops plus nasals and fricatives.

### 3.1.3.2 Place systems for each manner

Table III reports the content of place systems for each manner. It presents the spreading of plosives, nasals, fricatives and affricates according to the number of articulation places. We observe the following tendencies:

- There is neither a phonological system without plosives, nor a system with only one. When a system distributes the plosives across 2 places, these places are coronal and bilabial. Among 3 places, they are coronal, bilabial and velar. The glottal place appears in 4-places systems. Systems with 5 places recruit bilabial, coronal, velar, glottal and uvular consonants.
- If there is only one nasal in a system, it is a coronal one. If the system has 2 nasal consonants, they are bilabial and coronal. A velar nasal appears in 3 place systems. In larger systems (4 places), the palatal articulation emerges.
- When the fricatives of a system use only one articulation place, they are alveolar, i.e. coronal. When 2 places are used, they are coronal (alveolar) and glottal. In 3 place systems, labiodentals appear; in 4 place systems, the palatal area is exploited, and in 5 ones, velar fricatives emerge.
- Affricates use coronal articulation areas, whatever the place system size (from 1 to 3).

	0 place	1 place	2 places	3 places	4 places	5 places
<i>plosive</i>	0% -	0% -	0.2% coronal bilabial	31% coronal bilabial velar	<b>43%</b> <b>coronal</b> <b>bilabial</b> <b>velar</b> <b>glottal</b>	21% coronal bilabial velar glottal uvular
<i>nasal</i>	3.5% -	1.8% coronal	31% coronal bilabial	<b>32%</b> <b>coronal</b> <b>bilabial</b> <b>velar</b>	27% coronal bilabial velar palatal	3,5% coronal bilabial velar palatal coronal 2
<i>fricative</i>	7% -	6.6% coronal	18% coronal glottal	<b>28%</b> <b>coronal</b> <b>glottal</b> <b>labiodental</b>	20% coronal glottal labiodental coronal 2	11% coronal glottal labiodental coronal 2 velar
<i>affricate</i>	33% -	<b>42%</b> <b>coronal</b>	20% coronal coronal 2	4.9% coronal coronal 2 coronal 3	- -	- -

**Table III:** Consonant categories (plosives, nasals, fricatives, affricates) from UPSID<sup>451</sup> are classified according to the number of articulation places, with the percentage of corresponding languages (e.g., 31% of 451 languages present a plosive system spread on 3 places, etc.; 0 place means an absent category). The small differences observed in the percentages between Table II and Table III come from the number and the choice of languages between the two versions of UPSID. Only the dominant systems are noted under the percentage values.

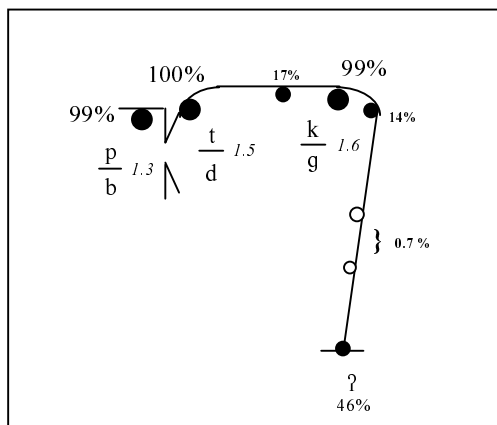
From these place system distributions per category, we can retain the following main tendencies:

- Generally, an n-element system contains a system with n–1 consonant places,
- When there is only one place, whatever the category (plosive, fricative...), coronal consonants are at stake,
- One can mention a possible link between the major tendencies of the place systems and the degree of articulatory “complexity” of the segments according to Maddieson (1984) and Laver (1994): “complex” segments (e.g. affricates, trills/taps/flaps) use almost exclusively the coronal area whereas “simple” segments use varied places.
- Grouping the most frequent manners, languages prefer contrast with 3 places: more than 1 out of 3 for plosives and nasals and almost 1 out of 3 for fricatives.
- A particular system size dominates each category: 43% of the languages have 4 places for the plosives and 74% have 3 or 4 places; 32% of the languages distribute their nasals on 3 places (59% on 3 and 4); 28% present 3 places for fricatives (48% on 3 and 4); 43% have only one articulation place for affricates.
- Combinations with 5 places exist in only 36% of the languages, mainly for oral plosives (97 languages).
- Combinations with 6, 7 or 8 places are not widely used (15% of the languages). We essentially find them in African systems.

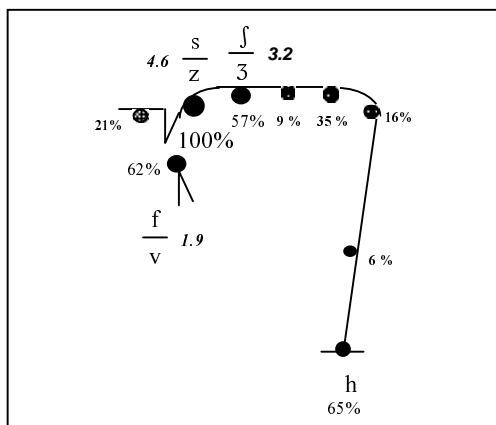
Our results (i) confirm that plosives are far and away the “star consonants”, ubiquitous in languages, and (ii) bring to light trends that appear very regular.

### 3.1.3.3 Voiced/voiceless ratio

Within the oral plosive category, voiceless sounds (64%) are a lot more numerous than voiced ones (Figure 9). Whatever the place of articulation, the voiceless feature concerns 72% of fricatives (Figure 10). Voiceless affricates are also dominant (74%).



**Figure 9:** Percentage of UPSID<sub>451</sub>’s languages which investigate the different places of articulation for oral plosives. The voiceless/voiced ratio is mentioned for the most frequent places of articulation. We observe a raising ratio from front to back places of articulation.



**Figure 10:** Percentage of UPSID<sub>451</sub>’s languages which investigate the different places of articulation for fricatives. The voiceless/voiced ratio is mentioned for the most frequent places of articulation. The predominance of voiceless fricatives is more important for alveolars and palatals.



### **3.2 Universal trends and ontogenesis**

We gathered the results of several studies on consonant content of production at different stages of canonical babbling (repetition of identical syllables) or variegated babbling (with different types of closures) – the latter being contemporary with the former (Vihman, 1996). All of these studies cover a period from 6-10 (Boysson-Bardies, 1996) to 15-24 months (Stoel-Gammon, 1985), for 15 languages including American English and French. Boysson-Bardies found that [p b m t d n k g ŋ] constitute 80% of consonant-like productions during the first months of babbling, [p b m t d] being the most frequent. The review of papers by Locke (1983) gives, by descending order of frequency, the inventory of consonants representing 95% of occurrences produced by children aged from 6 to 15 months of age: [b] and [m], [p], [d], [h] and [n], [t], [g] and [k], [j] and [w], [s]. Robb & Bleile's observations (1994) during a longer period (8 to 25 months of age) show that, for children aged from 8 to 12 months, several types of closures co-emerge. The most frequent productions are the plosives (oral and nasal) in alveolar and bilabial places. They thus confirm Davis & MacNeilage (1994), who showed that during 7-12 months of age, these consonant-like sounds cover 84% of infants' utterances ([t] [d] 46%, [n] 18%, [p] [b] 15%, [m] 5%); velars and glottals being the last to appear. Some of this experimental evidence has also been corroborated by simulations using linear midsagittal articulatory models developed for three different subjects: Vilain et al. (1999) found that pure mandibular raising, all other articulatory commands being frozen, leads to bi-labial, labio-dental or alveopalatal contacts, but never to velar contacts (also Abry et al., 1997).

If one draws a parallel between the results on language acquisition and typological data, one finds that the types of closures produced during the first months of babbling (8-12 months of age), as well as sounds acquired later during the second year, correspond to the consonant phonemes most frequently used in languages (apart from the order of emergence). One also finds that the preferred place systems, as shown for each manner in our typological analysis, are dominant in the production of babbling: alveolar and bilabial before velar and glottal in the plosive manner (oral or nasal); alveolar and glottal in the fricative manner; labiovelar, coronal and palatal in approximants ([l] and [j] being acquired later, at around 13-16 months of age (Robb & Bleile, 1994). All in all, we thus note that the consonant-like sounds produced during babbling's mandibular oscillations – the Frame, MacNeilage (1998) – correspond to the most frequent consonant inventories of the world's languages – in particular, they tightly correspond to languages with small inventories (Stefanuto, 1997). These findings allow us to form the hypothesis that the more frequent consonants in languages are, most probably, drawn from the stock of potential abilities of babbling. From an articulatory point of view, is it because they are produced with the greatest of ease? Locke & Pearson (1992, p. 26) did not hesitate to conclude: "*Infants heavily favor stop consonants over fricatives, and there are languages that have stops and no fricatives but no languages that exemplify the reverse pattern. [Such] "phonologically universal" patterns, which cut across languages and speakers are, in fact, the phonetic properties of Homo sapiens.*" Ontogenesis thus clearly contributes to supplying some determinating elements in the dependencies between certain general tendencies of phonological systems and the sensorimotor abilities of *Homo loquens*.

### 3.3 Hints towards explaining main tendencies

#### 3.3.1 Plosive vs. fricative consonants' predominance

Plosive consonants are characterised by the occurrence of a complete occlusion of the vocal tract somewhere between the glottis and the lips (Ladefoged & Maddieson, 1996; Crystal, 1997; Stevens, 1998). The main mechanism underlying the production of plosives is the generation of an impulsive acoustic source associated with the sudden release of the occlusion (Pelorson, 1997). From the control point of view, the main requirement for the articulators is thus to ensure a proper occlusion: this might be achieved by aiming at a virtual articulatory target that would be located beyond the actual vocal tract boundaries. For instance, in the case of the production of an apico-dental plosive, the tongue tip could be assigned a target that would be located slightly above the surface of the hard palate in the alveolar region: this would result in a full contact between tongue and palate, and thus ensure the complete occlusion of the vocal tract. This strategy has successfully been implemented for plosives (Bailly, 1997). It is important to note that this type of strategy does not require a precise control of the articulators: in a first approximation, a broadly defined target, as far as it is set beyond the physical limit of the involved vocal tract boundary, can ensure the realisation of the occlusion.

The situation of fricative consonants is quite different. Fricatives are characterised by the presence of a constriction somewhere in the vocal tract. This constriction induces a local acceleration of the air particle velocity that in turn induces, in conjunction with the shape of vocal tract, turbulence, and thus produces acoustic friction noise sources (Shadle, 1990; Badin et al., 1995; Stevens, 1998; Mawass et al., 2000). The generation of noise is precisely related to the size of the constriction, and in particular to its minimal constriction area. Specifically, if the constriction area is too small, the vocal tract will behave as in the case of plosives; on the other hand, if the constriction area is too wide, no friction noise will be generated and thus an approximant, or frictionless continuant, will be produced.

It appears thus clearly that articulatory control requires a much higher degree of precision for the production of fricatives than for the production of plosives. This fact constitutes indeed an important argument for the explanation of the preponderance of plosives over fricatives not only during the first months of babbling but also in the sound structures of the world's languages.

#### 3.3.2 Voiceless vs. voiced fricatives' predominance

Most fricatives and plosives in the world's languages can be voiced or voiceless. Before discussing issues about voicing, let us state again the basic principles underlying the generation of both voice and friction acoustic source for fricatives. The presence of voicing, i.e. of vocal fold vibration, is mainly controlled by an average pressure drop across the glottis and by vocal fold adduction (or glottis area at rest) (*cf. e.g.* Pelorson et al., 1994). In turn, the pressure drop across the glottis depends on both the glottis area and the geometry of the entire vocal tract. Note in passing that the control of the frequency of vocal folds oscillations is more specifically, but not only, associated with vocal fold length. Finally, the generation of friction is essentially governed by the oral constriction area and by the pressure drop across this constriction (Badin et al., 1995; Stevens, 1998).

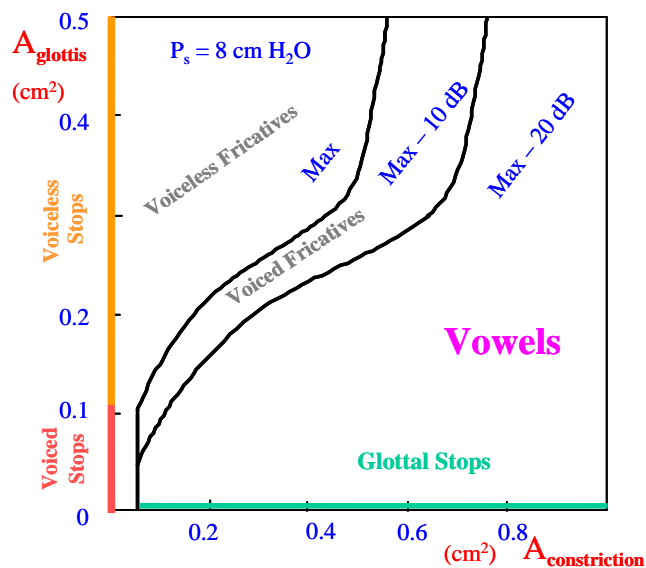
The coordination between glottis and oral constriction gestures thus plays a crucial role for acoustic excitation sources in the vocal tract, as has been widely recognised and discussed in the literature (e.g. Scully, 1971; McGowan et al., 1995; Badin et al., 1996; Stevens, 1998). From an aerodynamic perspective, the vocal tract can be viewed as two lumped constrictions between which the subglottal pressure is distributed (*cf. e.g.* Badin et al., 1996; Mawass et al., 2000). Although some refinements could be brought to this crude approximation (*cf.* Pelorson et al., 1994), it is expected that the discussion below will remain valid.

Denoting by  $A_g$  the low frequency component of the glottis area, and by  $A_c$  the oral tract minimum constriction area, the pressure drops at the glottis  $\Delta P_g$  and at the constriction  $\Delta P_c$  are given by the following equations:

$$P_s = \Delta P_g + \Delta P_c, \text{ with } \Delta P_c = \frac{\rho}{2} \frac{U^2}{A_c^2} \text{ and } \Delta P_g = \frac{\rho}{2} \frac{U^2}{A_g^2}$$

where  $P_s$  is the subglottal pressure,  $\rho$  the air density, and  $U$  the constant volume flow velocity. It is also known that the amplitude of voicing increases with  $\Delta P_g$  and reaches a maximum for a given  $A_g$  depending on  $\Delta P_g$  (Stevens, 1998), while the amplitude of the frication noise source is proportional to  $\Delta P_c P \cdot A_c^q$  (Badin et al., 1995).

It follows that, for a given subglottal pressure, simultaneously increasing voicing and frication amplitudes is contradictory: a balance must necessarily be found for voiced fricative consonants. Preliminary perceptual tests, where the ratio between voicing and frication noise amplitude was varied, showed that the consonants in synthetic vowel-fricative-vowel sequences were deemed acceptable voiced fricatives when the ratio was less than 12 dB (Mawass, 1997). Therefore, simulations performed with a complete vocal tract model, including a simplified aerodynamic model (Badin et al., 1996), were used to determine, for a simplified [ʃ] articulation, the region of the  $[A_g/A_c]$  control space where balance between voice and frication noise levels is reached within 12 dB (Mawass, 1997; Abry et al., 1998). The boundaries of this region, schematically drawn from the original simulation data, are shown in Figure 11. This region where the combinations of  $A_g$  and  $A_c$  values results in a balance between the voice and noise components is indeed very narrow: a rather strict coordination between the glottis and the oral constriction is therefore needed to produce acceptable voiced fricatives (*cf.* also Mawass et al., 2000). As already noticed by Ohala (1983) on qualitative grounds, this observation contributes to the explanation of the lower proportion of the voiced fricatives (Figure 10) in comparison with the voiceless ones in the world's languages (Stefanuto & Vallée, 1999).



**Figure 11:** General phase space for manners.

It shows a narrow region in the ( $A_g/A_c$ ) control space which ensures an approximate balance between voicing and frication noise for fricative consonants (simplified and adapted from real simulation data, Mawass, 1997; Abry et al., 1998).

## 4 Discussion and perspectives

A whole collection of data gathered during the course of the second half of the 20th century has progressively rendered untenable the principle of strict independence of form and substance. Languages do not construct their sound systems from “amorphous” materials, to quote a Saussurean expression. Entirely on the contrary, typological analyses carried out on phonological data that have been collected, standardised, and organised, indicate according to current evidence that languages, from whatever linguistic family, do not exploit the possibilities of the vocal tract, auditory and visual systems arbitrarily to organise their phonological structures (be it for the sign of a feature, the choice of that feature, the recruitment of phonemes, or the nature and organisation of syllables). Models that take production and perception constraints into account allow the prediction of broad tendencies and variants for vowel systems. The correspondence between the consonant phonemes that are most frequent in the languages of the world and the consonants that are most often produced during the first stage of babbling, whatever the language of the infant’s native environment, leads to propose the hypothesis that the consonant sounds of different languages are undoubtedly drawn from the stock of potential babbling capabilities.

The phonological typologies and universal tendencies that have been observed are unrelated to the linguistic families which served as the basis for the UPSID samples, currently the most representative database of sound structures. One of the interesting inferences that can be drawn from this lack of correlation is that the general tendencies of linguistic systems could depend on the phonetic substance of the sound inventories; it seems that it might be possible to find confirmation of this in ontogenesis, reinforcing the hypothesis that there might exist “phonetic properties of Homo Sapiens”. Of course, certain phonemes or features (rounding, length, nasality) are associated with certain families or rather with certain geographical areas (for instance /y/ in Indo-European and Uralo-Altaic families). This can only confirm that

typological classifications of sound structures and genetic classifications according to linguistic family are far from being identical, and that typology exhibits geographical tendencies. It is likely that statistical properties over the course of time, within a predetermined zone of the linguistic space, reflect the same constraints as the current general distribution, more or less, and that, as regards certain features, fluctuations of systems in neighbouring zones are produced more or less in phase.

Substance-based linguistics is an old dream for phoneticians, expressed in the most vigorous and brilliant terms more than twenty years ago by Bjorn Lindblom, until his recent formula: “derive language from non-language.” Vowel system prediction undoubtedly provides the most obvious success in this program. It is now possible to use a theory based on local and global perceptual (namely “non-linguistic”) arguments, a computational framework, and a quantitative methodology.

The Dispersion-Focalisation Theory (DFT) allows the prediction of vowel systems thanks to competition between two perceptual costs; the associated “phase spaces” determine the DFT winner in this space. We derive from the comparison between experimental phase spaces and UPSID data a region for which the theory predictions fit quite well with the phonological inventories. We have proposed an additional ingredient in our modelling phase-space framework, with the notion of polymorphism of a given phase, or superstructures in phonological organisations. Thanks to this new ingredient, we were able to obtain important results. A next step should consist in adding 8- and 9-vowels systems to the present analysis. Indeed, our analysis of raw UPSID<sub>451</sub> data demonstrate that more than 9 languages out of 10 (91.5%) in the database contain 3 to 9 vowel qualities. The difficulty however is that phase spaces become increasingly complex with such a large number of vowels.

But whatever the feasibility of this next step, we believe that the present results provide a new illustration of the convincing ability of substance-based theories, from Stevens' Quantal Theory and Lindblom's Dispersion Theory to ICP Dispersion-Focalisation Theory, to produce realistic predictions and useful typologies, at least for vowel systems.

The typology of consonant systems also shows that languages only use a small inventory of consonants even though they potentially dispose of a considerable number of possibilities. The place systems “preferred” by each manner show that languages' choices are at least partly conditioned by the morphological constraints that facilitate or complicate, make possible or impossible, the articulatory gestures according to the places, e.g. the complexity of the articulatory control of voiced and aerodynamics of fricatives. Although pleading the cause of a non-substantialist linguistics, Trubetzkoy conceded in 1939 (1970 p. 135): “*This fact cannot be ascribed to chance, and must be deeply rooted in the nature of the three series in question.*”

In fact, it is quite likely in the context of the syllable, associating consonants and vowels, that one must elaborate and test the models predicting general trends of sound structures. MacNeilage's theory, that he describes as a basic principle influencing the very structuring of speech itself, separates the *frame* level (syllabic) and the *content* one (segmental). The nature of mandibular *pure frames* in the canonical babbling stage is placed by MacNeilage in a labial frame. This hypothesis has been tested with 3 individual articulatory models (Vilain, 2000). We think it should rather be placed in a mandibular coronal frame coupled with a visual bilabial frame – which could explain the [d]/[b] coemergence in babbling. This is supported by arguments from articulatory simulations and by our typological arguments on the world's languages.

Almost thirty years after Lindblom's and Liljencrant's proposition, the viability of the project of a substance-oriented linguistics seems to be confirmed. Speech is by nature an interdisciplinary object of research, lying at the crossroads of several sensorimotor systems involved in the production and perception of biological communication signals, and of a major human competence, the faculty of language. Around this particular object of research have arisen groupings of several disciplines (ranging from acoustics, even fluid mechanics, to phonology and phonetics, while traversing biology, psychology, and information processing), placing speech at the centre of research that is as much fundamental as it is applied. All of the mutations that phonetics has undergone have provoked a reorganisation of the scientific connections between the phonetic sciences and adjacent disciplines. New interactions are clearly underway between linguistic science, cognitive science, and certain sectors of the physical and engineering sciences.

Hence the importance of "laboratory phonology" which tends to combine experimental phonetics, experimental psychology, and phonological theory (Ohala & Jaeger, 1986). This approach aims to subject hypotheses of phonological organisation to the kinds of validations used in the experimental sciences, which has been lacking until to date in generative phonology. At the end of this century, certain phonologists are thus clearly revising their positions: *"Beyond the differences which traverse it, the current phonological movement tends to give an objective status to the entities which are isolated by analysis. The question of substance, which has for so long been taken to be a secondary issue, only just likely to arouse the interest of certain empirical spirits (phoneticians), is the subject of a renewal of interest. [...] The anchoring of the reality of phonological systems in substance now appears to be an indispensable means for thinking of phonology as an interface between physical objects and phonological entities."* (*De natura sonorum*, Laks & Plénat, 1993). Our approach at the core of the relationships between phonetics and phonology shows that the perception-action interplay contribute to organise the sound structures of the world's languages. Phonetic knowledge can thus explore and make precise the natural constraints that all phonological theories must respect in order to satisfy concerns of (neuro)physiological plausibility. The reintegration of phonology into the natural order of things needs no longer involve a subordinate relationship between the two disciplines.

## References

- Abry, C., Badin, P., Mawass, K. & Pelorson, X. (1998) The Equilibrium Point Hypothesis and control spaced for relaxation movements or “When is movement actually needed to control movement ?”. *Bulletin de la Communication Parlée*, 4, 27-33.
- Abry, C., Badin, P., Vilain, A., Stefanuto, M. & Boë, L.-J. (1997) Une théorie de l'ontogenèse à l'épreuve de 3 modèles articulatoires anthropomorphiques. In *Un Générateur des Structures Sonores du Langage inspiré des contraintes développementales*. Rapport Final ARASSH (Agence Rhône-Alpes pour les Sciences Sociales et Humaines). Institut de la Communication Parlée, 1998.
- Alcantara, V. (1998) *Le trait de labialité dans les langues du monde. De la typologie des systèmes vocaliques à l'enseignement du FLE*. Travail d'Étude et de Recherche de Maîtrise, Université Stendhal, Grenoble.
- Badin, P., Mawass, K., Bailly, G., Vescovi, C., Beautemps, D., & Pelorson, X. (1996) Articulatory synthesis of fricative consonants: data and models. In *Proceedings of the 4th Speech Production Seminar – 1st ESCA Tutorial and Research Workshop on Speech Production Modeling: from Control Strategies to Acoustics*, Autrans, France, 221-224.
- Badin, P., Mawass, K. & Castelli, E. (1995) A model of frication noise source based on data from fricative consonants in vowel context. In Elenius, K. & Branderud, P. (Eds.), *13th International Congress of Phonetic Sciences ICPHS95*, Stockholm, Sweden: Arne Strömbergs Grafiska Press, 202-205.
- Bailly, G. (1997) Learning to speak. Sensori-motor control of speech movements. *Speech Communication*, 22 (2-3), 251-267.
- Boysson-Bardies de, B. (1996) *Comment la parole vient aux enfants*. Odile Jacob, Paris.
- Chomsky, N. (1965) *Aspects of the theory of syntax*. MIT press: Cambridge, Mass., USA. French translation: *Aspects de la théorie syntaxique*, Paris: Seuil, 1971.
- Chomsky, N. & Halle, M. (1968) *The sound pattern of english*. New York: Harper & Row. *Principes de phonologie générative*. French translation of parts I and IV by P. Encrevé, Paris: Seuil, 1973.
- Creissels, D. (1994) *Aperçus sur les structures phonologiques des langues négro-africaines*, 2<sup>nd</sup> ed., ELLUG, Grenoble, 322 p.
- Crothers, J. (1978) Typology and universals of vowel systems in phonology. In *Universals of Human Language* (2), Ed. by H.J. Greenberg, Stanford University Press, Stanford, California, 95-152.
- Crystal, D. (1997) *Dictionary of Linguistics and Phonetics*. Blackwell, Oxford.
- Davis, B.L. & MacNeilage, P.F. (1994) Organization of babbling: A case study. *Language and Speech*, 37 (4), 341-355.
- Ducrot, O. & Schaeffer, J.-M. (1995) *Nouveau dictionnaire encyclopédique des sciences du langage*. Paris: Seuil.
- Greenberg J., Ferguson C. & Moravcsik, E. (Ed.) (1978) *Universals of Human Languages: Method and Theory*, Phonology, Word Structure, Syntax, Stanford University Press.
- Hagège, C. (1982) *La structure des langues*. PUF, Paris.
- Hockett, C.F. (1955) *A Manual of Phonology*. Waverly Press, Baltimore.
- Jakobson, R., Fant, C.G.M. & Halle, M. (1952) *Preliminaries to Speech Analysis*. The MIT Press, Cambridge.
- Keating, P.A. (1990) Coronal places of articulation. *UCLA WPP*, 74, 35-60.
- Ladefoged, P. & Maddieson, I. (1996) *The sounds of world languages*. Blackwell Publishers, Oxford.
- Laks, B. & Plénat, M. (1993) *De natura sonorum*. Presses Universitaires de Vincennes, Vincennes.
- Laver, J. (1994) *Principles of Phonetics*. Cambridge University Press, Cambridge.
- Liljencrants, J. & Lindblom, B.E.F. (1972) Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language*, 48, 839-862.
- Lindblom, B. (1986) Phonetic universals in vowel systems. In *Experimental phonology*, J.J. Ohala (Ed.), New-York: Academic Press, 13-44.
- Lindblom, B. & Maddieson, I. (1988) Phonetic universals in consonant systems. In *Language, Speech and Mind. Studies in Honour of Victoria A. Fromkin*. L.M. Hyman & C.N. Li. (Eds.) Routledge, London and New-York, 62-78.
- Locke, J.L. (1983) *Phonological acquisition and change*. Academic Press, New-York.
- Locke, J.L. & Pearson, D.M. (1992) Vocal learning and the emergence of phonological capacity. A neurobiological approach. In *Phonological development. Models, Research, Implications*, C. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), Timonium, Maryland, 91-129.
- MacNeilage, P.F. (1998) The Frame/Content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21 (4), 499-511.
- MacNeilage, P.F. & Davis, B.L. (1990) Acquisition of speech production: Achievement of segmental independence. In *Speech Production and Speech Modeling*, W.I. Hardcastle & A. Marchal (Eds.), Kluwer, Dordrecht.
- Maddieson, I. (1984) *Patterns of sounds*. Cambridge studies in speech science and communication, Cambridge: Cambridge University Press, 2<sup>nd</sup> ed. 1986.
- Maddieson, I. & Precoda, K. (1989) Updating UPSID, *UCLA WPP*, 74, 104-111.
- Mawass, K. (1997) *Synthèse articulatoire des consonnes fricatives du français*. Thèse de doctorat, INP Grenoble.

- Mawass, K., Badin, P. & Bailly, G. (2000) Synthesis of French fricatives by audio-video to articulatory inversion. *Acta Acustica*, 86, 136-146.
- McGowan, R.S., Koenig, L.L. & Löfqvist, A. (1995) Vocal tract aerodynamics in [aCa] utterances: Simulations. *Speech Communication*, 16, 67-88.
- Ohala, J. (1983) The origin of sound patterns in vocal tract constraints. In *The Production of Speech*, by P.F. MacNeilage (Ed.), Springer Verlag, New-York, 189-216.
- Ohala, J. & Jaeger, J.J. (Eds.) (1986) *Experimental phonology*. Academic Press Inc., Orlando.
- Pelorson, X. (1997) On the fluid mechanics of bilabial plosives. *Speech Communication*, 22, 155-172.
- Pelorson, X., Hirschberg, A., Van Hassel, R.R., Wijnands, A.P.J. & Auregan, Y. (1994) Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation. Application to a modified two-mass model. *Journal of the Acoustical Society of America*, 96, 3416-3431.
- Prince, A. & Smolensky, P. (1993) *Optimality theory: constraint interaction in generative grammars*. ms, Rutgers University, Université du Colorado, New Brunswick et Boulder.
- Robb, M. & Bleile, K. (1994) Consonant inventories of young children from 8 to 25 months. *Clinical Linguistics and Phonetics*, 8, 295-320.
- Saussure, F. De (1916) *Cours de linguistique générale*. Publication de Ch. Bally & A. Sechehayé, Payot, Lausanne, Paris. Reedition by De Mauro, T., Payot, Paris, 1972).
- Scully, C. (1971) A comparison of [s] and [z] for an english speaker. *Language and Speech*, 14, 187-200.
- Schwartz, J.L., Abry, C., Boë, L.J., & Cathiard, M. (2002) Phonology in a theory of perception-for-action-control. In *Phonology : from Phonetics to Cognition*, Durand J. & B. Laks (Eds.), Oxford University Press, Oxford, 255-280.
- Schwartz, J.-L., Boë, L.-J., Vallée, N. & Abry, C. (1997a) Major trends in vowel system inventories. *J. of Phonetics*, 25 (3), 233-253.
- Schwartz, J.-L., Boë, L.-J., Vallée, N. & Abry, C. (1997b) The Dispersion-Focalization Theory of vowel systems. *J. of Phonetics* 25 (3), 255-286.
- Schwartz J.L., & Escudier, P. (1987). Does the human auditory system include large scale spectral integration?. In *The Psychophysics of Speech Perception*, M.E.H. Schouten (Ed.), Nato Asi Series, Martinus Nijhoff Publishers, Dordrecht, 284-292.
- Schwartz, J.L., & Escudier, P. (1989). A strong evidence for the existence of a large scale integrated spectral representation in vowel perception. *Speech Communication*, 8, 235-259.
- Sedlak, P. (1969) Typological considerations of vowel quality systems. *Working Papers on Language Universals*, 1, Stanford University, 1-40.
- Shadle, C. (1990) Articulatory-acoustic relationships in fricative consonants. In *Speech Production and Speech Modelling*, W. Hardcastle and A. Marchal (Eds.), Kluwer Academic Publisher, 187-209.
- Stefanuto, M. (1996) *Typologie des lieux d'articulation des langues du monde*. Travail d'Étude et de Recherche en Sciences du Langage, Université Stendhal, Grenoble.
- Stefanuto, M. (1997) *Apport de la neurocognition comparée, développementale et pathologique dans l'étude des tendances universelles phonologiques*. Mémoire de DEA en Sciences du Langage, Université Stendhal, Grenoble.
- Stefanuto, M. & Vallée, N. (1999) Consonant systems: From universal trends to ontogenesis. *Int. Congr. of Phonetic Sciences ICPHS99* (3), San Francisco, 1973-1976.
- Stevens, K.N. (1972) The quantal nature of speech: Evidence from articulatory-acoustic data. In *Human communication: A unified view*, P.B. Denes & J.R. Davis (Eds.), McGraw-Hill, New York, 51-66.
- Stevens, K.N. (1998) *Acoustic Phonetics*. The MIT Press: Cambridge, MA, London, England.
- Stoel-Gammon, C. (1985) Phonetic Inventories 15-24 months: A longitudinal study. *J. of Speech and Hearing Research*, 23, 506-512.
- Trubetzkoy, N.S. (1939) Grundzüge der Phonologie. *Travaux du Cercle Linguistique de Prague*, 7, 272 p. Translation by Cantineau, J. (1970) *Principes de phonologie*. Klincksieck, Paris, 394 p.
- Vallée, N. (1994) *Systèmes vocaliques: de la typologie aux prédictions*. Thèse de Doctorat en Sciences du Langage, Université Stendhal, Grenoble.
- Vallée, N., Schwartz, J.-L. & Escudier, P. (1999) Phase spaces of vowel systems. A typology in the light of the Dispersion-Focalization Theory (DFT). *Int. Congr. of Phonetic Sciences ICPHS99* (1), San Francisco, 333-336.
- Vihman, M.M. (1996) *Phonological development: The origin of language in the child*. Blackwell Publishers.
- Vilain, A. (2000) Apports de la modélisation des degrés de liberté articulatoires à l'étude de la coarticulation et du développement de la parole. Thèse de doctorat, Université Stendhal, Grenoble.
- Vilain, A., Abry, C., Brosda, S. & Badin, P. (1999) From idiosyncratic pure frames to variegated babbling: Evidence from articulatory modelling. *Int. Congr. of Phonetic Sciences ICPHS99* (4), San Francisco, 2497-3500.