Form and substance relationship in rhythmic structuring: a morphodynamic analysis of rate sensitivity at the infra-syllabic level

Valérie Pasdeloup°, Robert Espesser°, David Piotrowski□ & Malika Faraj*

*Université de Rennes 2, France
°Laboratoire Parole et Langage, UMR 6057 CNRS, Université de Provence, France
□Centre de Recherche en Epistémologie Appliquée, UMR 7656 CNRS, Ecole Polytechnique
valeriepasde@yahoo.fr, robert.espesser@lpl.univ-aix.fr,
david.piotrowski@polytechnique.fr, malika.faraj@free.fr

Abstract

The aim of this work is to determine in French, through an analysis of the rate sensitivity of vowels and consonants, how the speech rate is related to the prosodic structure at the infra-syllabic level. We present an analysis of the phonemic duration of a one thousand word speech corpus produced at three rates by one speaker with two repetitions. Results show that in French consonants as well as vowels are more rate-sensitive when stressed than when unstressed. This stronger rate sensitivity bears on all syllabic constituents. Still, consonants are less rate-sensitive than vowels and this is especially true for the stressed phonemes. We will account for the rate sensitivity of phonemes with a morphodynamic conception of phonology: where the acoustic substrate controls the phonemic form. We will see that the perception of speech rate is controlled by the same parameter which controls the features +/- stressed.

1. Introduction

1.1. Relationship between form and matter: theoretical background

The question of the relationships between the linguistic form and the substrate where it is actualized is one of the most foundational in linguistic science. The two most opposite conceptions of the connection between form and substance are found in the saussurian and the hjelmslevian theories [1, 2].

In saussurian structuralism, linguistic units are purely oppositive, negative and differential. Their meaning is essentially positional and, thus, they are linked to a geometrical and topological intuition of structures: the negative units result from the categorization of a substrate space by a net of boundaries. In this conception, the linguistic form has no existence outside the substrate space.

In the hjelmslevian conception, by contrast, the linguistic form is a pure abstract form which exists outside of any substrate. Using hjelmslevian terminology, we have to make a distinction between (i) (pure) “form”, (ii) “matter” where it can be instantiated by a process of “projection”, and (iii) the fusion of form and of “matter” which is named “substance”. Different forms can thus be applied to matter to produce different substances. For instance, the phonological, the articulatory or the phono-acoustic forms can differently structure the matter of expression.

The projection of form onto matter consists in assigning relational specifications of the form to the “single” units of matter. Therefore, the natural theoretical framework of such a conception of the relationship between form and matter is the theory of formal systems. It is thus clear that in this conception matter is unable to have any effect on linguistic structure; first because it is itself amorphous, and second, because form is completely defined and built outside it.

On the other side, the theoretical background of the saussurian conception and its mathematical explanation are given by the morphodynamic approach (catastrophe theory [3]) which gives an account of the categorization of a substrate set (acoustic, semantic...) by emergence of a net of boundaries. Precisely, “the categorization is the trace on the control space [or substrate] of the instabilities and conflicts of the internal states it controls [and which are reached through dynamic processes]” [4, p. 96]. Thus here the form is not independent of the substrate: it is instantiated in the substrate which has a control function on it. The morphodynamic approach has been successfully used to explain various categorization schemes in speech and semantics, such as the qualitative and privative phonemic oppositions, or the semiotic square [4, 5].

It is important to underline that morphodynamic complexity is neither free nor arbitrary. The net of boundaries that categorizes a substrate space to institute differential identities is highly constrained by (i) a principle of stability and (ii) by the dimension N of the control set. In the basic case (when dynamics are potential functions) the determination and the classification of all possible differential structures has been achieved (Thom) for $N < 5$. One methodological consequence is that, for a given number N of control parameters, the empirical observation of differential organization and the distribution of structural identities which can be put in correspondence with the morphodynamic structure determined by this number N, has to be seen as a proof in favour of this approach and, consequently, in favour of the existence of a control relationship between substrate and form.

In this paper, we will establish the rate sensitivity of phonemes to bring evidence in favour of a morphodynamic conception of phonology: where the acoustic substrate controls the phonemic form.

1.2. Requirements of matter in rhythmic structuring

1.2.1. Speech production as a disembodied object

In the phonological structure generated by standard intonation theories, motor, perceptual and cognitive constraints which govern other human productions and their percepts are generally not taken into account. Speech productions are often considered as pure, disembodied, formal objects without any size, weight or mass. Their concrete realization in substance is not considered in their formal representation since the representation of time and duration is usually missing. Time representation only appears in the linear representation of linguistic constituants placed before or after another one. In contemporary syntactic theories for example, the time is apprehended through the reduction of highly hierarchized constituents in the linearity of the surface structure. It thus seems that there is a confusion between the linear ordering of terminal constituents of a syntagmatical structure and the requisite dimension of the form of time. In a paradoxical way, the dimension of time glides into the “dimension of the represented structure”, thus into a structural fact without any temporal dimension, and by doing so, supresses itself. From this point of view, the treatment of temporal dimension of linguistic phenomena reaches a dead lock.

The fact that contemporary syntactic theories handle disembodied linguistic objects does not seem to have major repercussions on their development. On the other hand, it is paradoxical to approach rhythmic and prosodic structuring from the point of view of the organization of disembodied linguistic objects: indeed how can we work on the actualization of linguistic objects in the very flow of their production if we neglect their actual material charateristics? This attitude is explained by the fact that linguistics had to evict speech matter in order to constitute itself as a science. Thus linguistics has pushed away matter in the acoustic and articulatory measurement techniques of the substance, reducing the materiality of linguistic objects to the measurement and labelling of their substance, that is reducing it to their representations.

1.2.2. Requirements of matter in speech: size, number and elasticity

Considering that rhythm integrates, in a unique acoustic substrate, matter requirements and form actualization, the matter/substance/form distinction becomes a key factor to understand rhythmic structuring. In previous works, we have shown for French that there are requirements on the size of rhythmic templates, such as stress group and rhythmic word [6, 7]. Therefore two sentences having the same syntactic structure but composed of a different number of syllables will not be given the same prosodic structure. In recent work, we show for French that there are different requirements on the speed of production of stressed and unstressed syllables [8]. Rhythm is not elastic: temporal structuring produced at a slow rate is not the consequence of a linear decrease of the same material pronounced at a fast rate. When speech rate changes, syllabic duration does not vary the same way whether the syllable is stressed or not.

1.3. Aim of the experiment

This experimental study is part of a project on requirements of matter in rhythmic templates and stress pulsation in French. The strategy we adopted is to constrain the rhythmic structure of read texts by manipulating speech rate. It is thus possible to observe requirements of matter while the same formal (lexical and morphosyntactic) linguistic structure is retained. The purpose here is to determine in French how the speech rate is related to the prosodic structure at the infra-syllabic level. Are vowels more rate-sensitive than consonants? More generally, is the relative temporal progress of the consonant and vowel affected by the speed of production?

We present an analysis of the phonemic duration of a one thousand word speech corpus produced at three rates by one speaker with two repetitions.

2. Method

2.1. Speech materials

The corpus is a one thousand word tale, produced at three rates (normal, fast and slow) by one speaker (the first author) with two repetitions and recorded in a sound treated recording booth. The best repetition was been selected for each rate. There were thus about 1200 syllables for each rate and 8081 phonemes for the three rates: 2660 at fast rate, 2698 at normal rate and 2723 at slow rate. Among the 8081 phonemes, there were: 4085 consonants, 3527 vowels (whose 423 schwas constitute a vocalic nucleus), 108 extrametrical schwas (which can not constitute the vocalic nucleus of a syllable, generally in a pre-pause position) and 361 semi-vowels.

2.2. Experimental analysis

The rhythmic structuring study of the corpus includes the phonetic analysis of prosodic parameters (mainly phoneme and syllable duration, pitch contour and pause) and their phonological interpretation, which allows the determination of an abstract rhythmic structure in the framework of a given theoretical model. Phonological representation corresponds to accentuation and rhythmic phrasing. Our rhythmic law-based prosodic model distinguishes four prosodic levels [7, 8]: the syllable (the minimal rhythmic unit which can be stressed or unstressed), the stress group (in French, this consists of one stressed syllable preceded by zero or a few unstressed syllables), the rhythmic word (the smallest prosodic structure which organizes a meaning group [9]) and the rhythmic sequence (the major prosodic structure).

The first step was to process the phonetic labelling of the corpus utterances and their phonetic alignment using the aligner developped by LORIA (Foher & Laprie: <http://www.loria.fr/equipes/parole/>). We manually corrected the labelling and the phonetic alignment. The program codes identically the oral and nasal vowels with two vowel qualities corresponding to the archiphonemes: /E Œ O A E~/. There were thus 17 consonant types and 11 vowel types, excluding the extrametrical schwas (which were not take into account in the phoneme analysis). Syllabification was processed by a Praat script and corrected manually. The next step was the phonetic analysis of prosodic parameters. It was carried out by the first and fourth authors. Consequently, the stressed or unstressed status of the syllables was determined in terms of accentual contrasts actually produced and perceived and not in terms of predictions of a formal grammar. In the last step, these data are interpreted in the framework of our model. The model specifies the categories of accent (primary or secondary) and the categories of rhythmic groups. In this research, only the accent interpretation was taken into account, that is to say the stressed or unstressed status of the syllables.

3. Results

3.1. Articulation rate and phoneme duration

The articulation rate was 15.31 phonemes/s at fast rate, 12.33 phon/s at normal rate et 9.88 phon/s at slow rate (extrametrical schwas and semi-vowels included, pauses excluded). For syllables, the articulation rate was 6.8 syll/s at fast rate, 5.4 syll/s at normal rate and 4.4 syll/s à slow rate. Table 1 shows that stressed (S) phonemes are more rate-sensitive than unstressed (U) ones. This phenomenon seems stronger in stressed vowels (V) which are more rate-sensitive than stressed consonants (C): compared to normal rate, stressed Vs are on average 34% shorter at fast rate and 39% longer at slow rate, while stressed Cs are on average 17% shorter at fast rate and 25% at slow rate. Unstressed Cs and Vs show a very similar rate insensitivity: 18ms mean difference between fast and slow rate conditions for unstressed Cs and 23ms for unstressed Vs. The greater variation of duration between C and V is observed at slow rate for the stressed ones.

Table 1 : Mean duration of unstressed and stressed consonants and vowels under the three rate conditions

	Fast rate	Normal rate	Slow rate	mean
Unstressed C	62.1ms ±25	68.4ms ±28	79.6ms ±33	70
Stressed C	79.1ms ±34	95.4ms ±41	119.1ms ±51	97.8
Unsressed V	57.8ms ±18	68.1ms ±21	80.6ms ±26	68.8
Stressed V	69ms ±24	104.2ms ±39	144.6ms ±75	105.9

3.2. Statistical analysis

Phonemic duration was analyzed in terms of three factors: *Rate* as a factor with 3 ordered levels (Fast, Normal, Slow), Stress as a 2 level factor (Unstressed, Stressed), *Class* as a 2 level factor (Consonant, Vowel). A mixed linear model, where phonem was the grouping factor, took into account the repetition of the 28 unbalanced phoneme groups (17 consonants and 11 vowels) ([10], <http://www.R-project.org/>). Consequently, the variations of interphonemic duration were neutralized. Moreover, the use of the logarithm of the phonemic duration has stabilized the variance. This first model showed that only the linear components of rate are significant. Therefore, the rate factor was treated as a classical numerical variable, which simplifies the model. Each rate was associated with the corresponding total duration of the corpus (without pauses). Then the *Rate* variable was centred on the Fast rate in order to test some hypothesis for this rate.

Table 2 shows that all interaction coefficients with *Rate* are significant and positive: *Rate:StressS*, *Rate:CLassV*, *Rate: StressS:ClasseV*. We thus obtain four distinct regression lines for Unstressed C, Unstressed V, Stressed C and Stressed V (Fig. 1). The significance of other coefficients is only meaningful for the Fast rate *StressS* (significant), *ClasseV* (not significant) and *StressS:ClasseV* (not significant).

Table 2 : Regressors of the centred mixed model

	Value	Std.Error	DF	t-value	p-value
(Intercept)	4.1308	0.0612	7577	67.47	0.0000
Rate	0.0025	0.0002	7577	14.54	0.0000
StressS	0.1556	0.0164	7577	9.48	0.0000
ClasseV	-0.0894	0.0973	26	-0.92	0.3667
Rate:StressS	0.0016	0.0002	7577	6.27	0.0000
Rate:ClassV	0.0008	0.0002	7577	3.47	0.0005
StressS:ClasseV	0.0108	0.0244	7577	0.44	0.6583
Rate:StressS:ClasseV	0.0019	0.0004	7577	5.30	0.0000

3.3. Interpretation

The significant interaction *Rate:StressS* shows that the unstressed phonemes are less rate sensitive than stressed ones (Table 1, Fig. 1). This finding confirms results showing for French a greater rate sensitivity of final syllables compared with penultimate ones [11]. The *Rate:ClassV* interaction shows a stronger rate effect for V than for C. The lesser elasticity of C was observed in previous works for French [12, 13] and others languages. The double interaction *Rate:StressS: ClassV* specifies that stress enhances the rate sensitivity of V. Consequently, V are more rate sensitive than C: this is especially the case for the stressed ones and to a lesser but significant proportion for the unstressed ones.

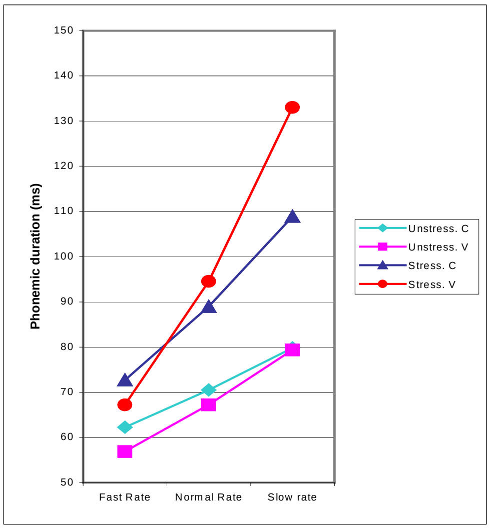

Figure 1 : *Estimated C and V duration by the mixed model under the three rate conditions*

Concerning the accentual contrasts of duration between the stressed phonemes and the unstressed ones, contrats strengthen when rate decreases and are significantly stronger for V than for C at Normal and Slow rates. At Fast rate, the contrast is maintained (*StressS*: significant) but does not differ significantly between V and C (*StressS:ClasseV*: not significant). In our opinion, this difference of rate sensitivity between C and V would correspond to universal constraints of matter (motor constraints of articulatory control and proprioceptive and auditory sensorymotor constraints). As a matter of fact, since Cs are intrinsically transitory phenomenon, when rate decreases they cannot lengthen beyond a certain limit (celling effect). On the contrary, the difference of rate sensitivity between stressed and unstressed phonemes would be phonological, that is determined by *form constraints* which can be explained in terms of control of a substrate space.

To give a morphodynamic explanation of these experimental results, we have first to notice that although the unstressed phonemes are rate sensitive, this variation is smaller than the one of stressed phonemes (the interaction coefficient *Rate:StressS* is positive and significant). Second, we will accept the assumption that there is a dual relationship between production and perception: when speaking +/- Fast, we produce phonematic structures which are dually perceived as +/- Fast speech. Then, assuming that the perception of speech rate is not local but global (i.e. the speech rate is not perceived on each phoneme but on a upper whole), it follows that the information of speech rate may be preferentially located in stressed phonemes. Furthermore, if we make the assumption that the smaller rate sensitivity of unstressed phonemes corresponds to an intrinsic (substrate) low level variation, then the smaller rate sensitivity of unstressed phonemes would not concern nor affect the global perception of speech rate.

From a morphodynamic point of view, this means that the perceptual dimension of speech rate is correlated to the phonemic opposition stressed/unstressed, which is controlled by the acoustic parameters of stress. The morphodynamic structure which accounts for such a case is a connection of two dynamic structures of "neutralization" (privative oppositions: +/- stressed and +/- perception of rate).

Let us consider first the parameter *s* defined as a function of acoustic parameters of stress. The parameter *s* controls the opposition of features +/- stressed (Fig. 2). For *s* greater than a value H, the state reached by the phonemic system is the minimum of the potential function (element of an internal space F of dynamics [1, 4, 5]) it controls: the state A (stressed), which is in (privative) opposition with the absence of state (unstressed) of the dynamic process controlled by the values *s* smaller than H.

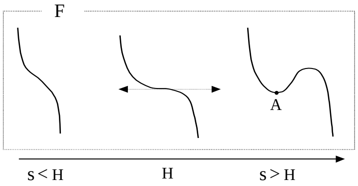

Figure 2: *Control of stress feature*

Let us consider now another internal space F' (of dynamical process), on the same control space, and which determines the actualization of the *linear dimension R of linguistic perception of rate*, that is the form by which acoustic data is perceived as linguistic duration (Fig. 3). In this case, for *s* greater than a value H (stressed phonemes), the phonemic system actualizes a particular value of rate R_i on the R dimension, and for *s* smaller than H, no state is actualized, which means that the linguistic perception of duration is neutralized.

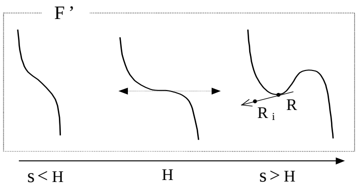

Figure 3: *Control of speech rate*

From a dual point of view, that is precisely what is observed in this experiment: the linguistic evaluation of rate is not called up for unstressed phonemes, which consequently leave their duration practically unchanged when the subject speaks more or less quickly. Fundamentally, it means that the perception of speech rate is controlled by the same parameter which controls the features +/- stressed.

4. Discussion

This study showed that, in French at the infra-syllabic level, consonants as well as vowels are more rate-sensitive when stressed than when unstressed. This stronger rate sensitivity, which we observed for stressed syllables [9, 10], bears on all syllabic constituents. Still, consonants are less rate-sensitive than vowels and this is especially true for the stressed phonemes. These results gives evidence that stress-bearing unit is the syllable (theory of Hayes [14] and Selkirk [15]) rather than the syllabic nucleus (Halle and Vergnaud [16]).

Concerning the motor programming, the results support the hypothesis that only the duration of stressed phonemes will be planned (that is to say the duration contrast). The duration of unstressed phonemes will not be planned. Therefore, the strong rate sensitivity of stressed phonemes would correspond to a high level variation of the system (i.e. motor commands) while the very weak rate sensitivity of unstressed phonemes would correspond to an intrinsic low level variation. Concerning the relationship between phonological markedness of stress and phonetic variation of rate, rhythmically unmarked (i.e. unstressed) phonemes are less sensitive to speech rate than rhythmically marked (i.e. stressed) phonemes.

To conclude, the linguistic constituents size (number of syllables of lexical words and phrases) as well as their speed of production determines the rhythmic flow structuring. Therefore requirements of matter occur during the rhythmic formatting on the size and on the number of linguistic forms as on their relative temporal progress inside and between linguistic sub-components. The morphodynamic structure gives us a theoretical account of this configuration (fig. 3). In summary, we have known for a long time now that linguistic forms determine speech matter to be a given substance rather than another. It seems reasonable now to consider that speech matter has a functional (control) effect on the linguistic form.

THANKS : We wish to thank "Master Daniel" Hirst for his helpful comments and useful scripts running on Praat.

5. References

[1] Piotrowski, D., 1997. *Dynamiques et structures en langue*, Éditions du CNRS, Paris.

[2] Hjemslev, L., 1943. Prolégomènes à une théorie du langage, trad. française : Les Editions de Minuit 1971.

[3] Thom, R., 1972. *Stabilité structurelle et morphogénèse*, W.A. Benjamin.

[4] Petitot, J., 2004 (1985), *Morphogenesis of meaning*, P. Lang, Bern.

[5] Petitot, J., 1992. *Physique du sens*, Editions du CNRS, Paris.

[6] Pasdeloup, V., 1990. *Modèle de règles rythmiques du français appliqué à la synthèse de la parole*, Thèse de doctorat, Université de Provence.

[7] Pasdeloup, V., 1992. A prosodic model for French text-to-speech synthesis: A psycholinguistic approach, In *Talking Machines: Theories, Models and* Designs, G. Bailly, C. Benoît & T.R. Sawallis (eds.), Elsevier Science Publisher.

[8] Pasdeloup, V. ; Espesser, R. ; M. Faraj., M., 2006. Rate sensitivity of syllable in French: a perceptual illusion? *3rd International Conference on Speech Prosody*.

[9] Vaissière, J., 1980. La structuration acoustique de la phrase française, In *Annali della Scuola Normale Superiore di Pisa* 3(10), 529-560.

[10] Pinhero, J. C. & Bates, D. M., 2001. *Mixed-Effects Models in S and S-Plus*, Springer.

[11] Duez, D., 1999. Effects of articulation rate on duration in read French speech, *Proc. Of Eurospeech*, Budapest.

[12] Fletcher, J., 1987. Some micro and macro effects of tempo change on timing in French, *Linguistics*, **25**, 951-967.

[13] Fletcher, J., 1991. Rhythm and final lengthening in French, *Journal of Phonetics*, **19**, 193-212.

[14] Hayes, B., 1981. *A metrical theory of stress rules*, Doctoral dissertation, MIT, distributed by IULC.

[15] Selkirk, E. O., 1984. *Phonology and Syntax*, Cambridge: The MIT Press.

[16] Halle, M. ; Vergnaud, J.-R., 1987. *An essay on stress*, Cambridge: The MIT Press.