LES INVARIANTS RELATIFS : OCCLUSIVES ET VOYELLES DIFFUSES

P. Durand

► To cite this version:

P. Durand. LES INVARIANTS RELATIFS : OCCLUSIVES ET VOYELLES DIFFUSES. Journal de Physique Colloques, 1990, 51 (C2), pp.C2-519-C2-522. 10.1051/jphyscol:19902122 . jpa-00230416

HAL Id: jpa-00230416

<https://hal.science/jpa-00230416>

Submitted on 1 Jan 1990

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LES INVARIANTS RELATIFS : OCCLUSIVES ET VOYELLES DIFFUSES

P. DURAND

*Institut de Phonétique, CNRS URA 256, Université de Provence
Aix-Marseille 1, Avenue R. Schuman, F-13621 Aix en Provence, France*

Résumé - Le but de cette étude est de mettre en évidence les problèmes posés dans les travaux sur la parole, la démarche qui consiste à mettre en parallèle deux types d'unités différentes, un paradigme de données normalisées et l'actualisation, sur l'axe syntagmatique d'unités de parole. L'étude de logatomes ($C_1V_1C_2V_2C_3$, avec C = Consonne occlusive sourde et V = /i,e,y ou ø/) et de données recueillies antérieurement montre la difficulté de concilier une invariance paradigmatique et la pertinence syntagmatique. La solution semble être dans la linéarité du signal, c'est à dire dans la recherche d'indices suffisants au niveau de l'énoncé et non à celui d'indices nécessaires au niveau paradigmatique.

Abstract - The aim of this paper is to show evidence that, in speech, the fact of comparing two kinds of different units, a paradigm of normalised data and the realization of utterances on the syntagmatic axis, raises obvious difficulties. The study of non-speech words ($C_1V_1C_2V_2C_3$, with C = unvoiced stop and V = /i,e,y or ø/) and previous datas show that paradigmatic invariance can hardly agree with syntagmatic relevance. At the utterance level, sufficient clues are to be searched in the linear signal and not in the selection of necessary clues on the paradigmatic level.

1. INTRODUCTION

Ce n'est qu'à partir de l'apparition de Visible Speech et du Sound Spectrograph⁽¹⁾ que put être entreprise de manière systématique l'étude de la parole continue. Au delà de certains insuccès, dus tant à des limitations techniques - artefacts - qu'à des conclusions hâtives⁽²⁾ ce sont surtout les réussites certaines auxquelles ont conduit ces études qui ont retenu l'attention. Cependant, au terme d'années de recherche, apparaissent un certain nombre de problèmes, dont la solution, obtenue parfois à l'aide de règles "ad hoc" montre assez qu'ils tiennent leur origine du cadre théorique dans lequel se sont insérées les connaissances acquises sur la substance acoustique de la parole continue. C'est dans cette perspective que se situe notre réflexion, qui va tenter de montrer son bien fondé en s'appuyant sur l'étude de logatomes combinant les occlusives sourdes et les voyelles /i,e,y et ø/ et sur des résultats obtenus antérieurement⁽³⁾.

2. LES ANALYSES DE LA SUBSTANCE PHONIQUE

Acquérir de nouvelles connaissances à l'aide de nouveaux procédés d'investigation, comme l'était la spectrographie, consiste à étudier comment se manifestent certains phénomènes déjà connus, au moyen de cet outil. Aussi, dans la mesure où les objectifs que se propose une science ne sont pas immuables, n'est-il pas indifférent de savoir comment s'orientaient les recherches de type linguistique au cours de cette période, puisqu'explicitement ou non, elles allaient sous-tendre les

(1) en particulier : POTTER, R.K., KOPP, G.A., GREEN, H., *Visible Speech*, New York, Dover Public. Inc, 1947 et KOENIG, W.H., DUNN, H.K., LACEY, L.Y., *The Sound Spectrograph*, J.A.S.A., 10/1, 1946, p.21-32;

(2) Un exemple parmi d'autres est donné par la critique de la notion de "spectrophonème" par Stevens, 1960, C.R. de Pulgram, 1959)

(3) DURAND, P., *Étude acoustique des consonnes occlusives du français commun*, Thèse dactyl. Université d'Aix-Marseille 1, 1982; DURAND, P., *Variabilité acoustique et invariance en français: consonnes occlusives et voyelles*, Paris, Éditions du C.N.R.S., 1985:

analyses acoustiques. De peu d'importance immédiate semble avoir été, l'opinion soutenue en 1947 par Hjelmslev⁽⁴⁾ selon laquelle le phonème, entité négative, est "une forme indépendante de toute substance d'expression" et donc, en conséquence que "l'analyse perdra nécessairement de sa rigueur dans la mesure où elle fera intervenir la substance phonique..."⁽⁵⁾ Plus proche des préoccupations des acousticiens de la parole semblent avoir été d'une part, les travaux des linguistes américains qui à la suite de Bloomfield se sont intéressés aux unités linguistiques, en particulier Pike⁽⁶⁾ et Harris⁽⁷⁾, et de l'autre et plus directement peut être encore, ceux de Jakobson et al⁽⁸⁾ qui tentaient de donner à la phonologie pragoise une généralisation de type acoustique.

Plus profondément la conviction de Bloomfield que deux énoncés ne sont jamais phonétiquement tout à fait identiques et que l'identification faite par l'auditeur se fait seulement à partir d'un petit nombre de propriétés retenues par la perception alors que toutes les autres sont indifférentes allait conduire à une recherche systématique de la simplification du signal, réduit aux seuls "distinctive features" qui s'y retrouvent toujours. L'étude des documents spectrographiques allait renforcer cette tendance. Ainsi que le souligne Fant⁽⁹⁾:

"Spectrographic picture convey an overflow of data which are not essential for descriptive purposes. This redundancy is in part a matter of interrelation, repetitions and continuities within the signal structure, in part the presence of a fine structure the details of which carry very little or no information."

Comme nous venons de le voir, l'analyse acoustique allait être sous tendue par un certain nombre de principes, et c'est à travers de ce moule contraignant qu'allait se poursuivre l'étude des éléments constituant la chaîne sonore. Si l'identification de la forme linguistique ne pouvait se faire qu'à travers des éléments toujours présents dans la représentation d'une unité de la chaîne sonore, c'est ces éléments qu'il fallait rechercher. Puisque l'analyse phonologique, au travers des Preliminaries parvenait à une description de toute réalisation phonétique au moyen d'un nombre fini de traits acoustiques, ce sont les indices correspondant à ces traits qu'il fallait isoler d'une masse amorphe d'éléments peu pertinents ou redondants. La vérification des indices découverts en analyse au moyen de la synthèse allait confirmer d'une certaine façon cette manière d'aborder le problème puisque au terme d'essais et d'erreurs elle mettait fatalement le doigt sur un élément du spectre contribuant à la perception de telle ou telle propriété du signal. Mais cette dernière technique même allait elle aussi être contrainte de faire appel à des recettes 'ad hoc' dès lors que l'on voulut parvenir à une synthèse de qualité. Les difficultés furent d'un autre ordre en ce qui concerne les techniques de reconnaissance de la parole. Au lieu d'avoir affaire à des auditeurs, dont on connaît la plasticité et la polyvalence du système de reconnaissance⁽¹⁰⁾ c'est à la multiplicité des stratégies phonatoires que le système va se trouver confronté. Et, au delà d'un certain pourcentage, parfois non négligeable de formes reconnues, avec un vocabulaire fermé ou un nombre de locuteurs limité se poseront des problèmes importants pour des systèmes totalement ouverts.

(4) Hjelmslev, L. Structural Analysis of Language, Studia Linguistica, I, 1947, pp.69-78 et Essais linguistiques, 1959, pp.27-35

(5) MARTINET, A. La linguistique synchronique, 1965, op. cit. p.126

(6) PIKE, K., Phonetics, Ann Arbor: The University of Michigan Press, 1943

(7) HARRIS, Z.S., Methods in Structural Linguistics, Chicago, University of Chicago Press, 1951

(8) JAKOBSON, R., FANT, G., HALLE, M., Preliminaries to Speech Analysis The Distinctive Features and their Correlates, Cambridge, The M.I.T. Press, 1951

(9) FANT, G., Acoustics of Speech, Proc. 11ème Congrès International d'Acoustique, Amsterdam, Elsevier, 1961 et Speech Sounds and Features, Cambridge, The M.I.T. Press, 1973, pp.3-16, op.cit. p.5

(10) Même des énoncés ayant subi une substitution de type phonologique seront considérés comme 'normaux' par un certain nombre de sujets pour peu que l'énoncé conserve une certaine cohérence syntagmatique. cf DURAND, P. Conflits de systèmes phonologiques Perception de groupes Consonnes-voyelles en français chez les étudiants marocains: étude préliminaire, à paraître TIPA, XII;

3. CRITIQUE MÉTHODOLOGIQUE.

Il est vrai que dès 1953, C.M. Harris⁽¹¹⁾ avait montré que l'onde de parole n'était pas réductible à la succession de segments sonores discrets, et que G.Fant⁽¹²⁾ avait montré les diverses conceptions possibles de la concaténation de l'énoncé. Il n'en reste pas moins que l'investigation phonétique se consacra, après la segmentation de l'énoncé, à la recherche des éléments présents dans chaque réalisation d'une unité de forme, à la recherche de valeurs centrales moyennes pour les caractériser. Cette manière de procéder montre bien le poids de la conception Bloomfieldienne du langage pour qui une réalisation sonore particulière est la réalisation d'un schème, ici acoustique, invariant que l'on doit dégager d'une gangue d' "accidents sonores", individuels et particuliers, sans aucune valeur du point de vue linguistique.

Il est possible, nous semble-t-il de comprendre en quoi la méthodologie suivie a conduit à une situation telle que le recours à des règles 'ad hoc'⁽¹³⁾ soit envisagé de préférence à une application stricte des principes théoriques. Toutes les analyses reposent sur la conviction que l'unité de type phonématique est par nature peu variable, que les variations que l'on peut enregistrer ne sont qu'accidentelles et limitées. Cette conviction a pour conséquence que seule importe la portion du spectre du son considéré où celui-ci est le moins susceptible de variation et que la valeur qui le caractérise sera la moyenne du plus grand nombre possible de mesures sur cette portion. Il va de soi que, fondée sur un très grand nombre de mesures centrales, ce procédé sera à même de rendre compte d'une grande majorité de réalisations de ce même son. Mais toutes les réalisations marginales seront justifiables de règles spécifiques.

Par contre, si nous nous plaçons à un autre niveau, c'est à dire à celui de l'énoncé, nous constatons que les rapports entre unités sonores ne sont pas de type paradigmatique, mais de type syntagmatique. C'est à dire que l'énoncé ne contraint pas cette unité phonique à posséder toutes les propriétés qu'elle doit avoir au niveau phonématique pour s'opposer aux autres unités de même rang du système. Elle ne doit avoir que celles qui lui permettent de s'opposer aux autres unités de même rang qui peuvent apparaître à sa place dans la même chaîne sonore ainsi qu'avec les sons qui l'entourent. Elle peut aussi acquérir, du fait du contexte où elle se trouve insérée, des propriétés nouvelles, à la limite contradictoires avec celles de l'unité de rang phonématique qu'elle représente⁽¹⁴⁾, du fait des contraintes articulatoires et communicatives qui peuvent lui être imposées en fonction de la taille des unités de programmation motrice.

Ainsi, la recherche d'invariants, telle qu'elle est couramment pratiquée, ne peut servir, au niveau des unités sonores qu'à caractériser imparfaitement les réalisations les plus 'moyennes' de ces unités sonores et encore souvent en niant comme non-pertinent des éléments présents dans le signal tout en regrettant la faiblesse ou l'absence d'éléments pertinents qui devraient s'y trouver.

4. POUR LA RECHERCHE D'INDICES SUFFISANTS

C'est par ce terme que nous désignerons les indices servant à caractériser une unité sonore dans une chaîne phonique donnée. Dans le cadre qui nous est imparti, et pour montrer le bien fondé de notre analyse, nous nous contenterons d'examiner comment, en syllabe initiale de logatomes disyllabiques les réalisations de [pi] ont des indices qui permettent de les différencier de toute syllabe

(11) HARRIS, C.M., A Study of the Building Blocks in Speech, *J.A.S.A.*, 25/5, 1953, pp.962-969;

(12) FANT, G., Descriptive Analysis of Acoustic Aspects of Speech, *Logos*, V, 1962, pp.3-17 et FANT, G., 1973, pp.17-31;

(13) L'utilisation de ces règles était souvent "légitimée" dans certains procédés de synthèse, comme le "Pattern Playback" par exemple, dans la mesure où elles permettaient d'obtenir des séquences identifiables même si elles étaient en contradiction avec le principe du "lecteur de sonagramme" de cet appareil, et si on a pu en tirer des conclusions erronées sur la nature des indices.

(14) Pour ne donner qu'un exemple, prenons la réalisation de [a] de maman et son caractère non-nasal dans le spectre.

commençant par [p] (ici, avec [e,y,ø] qui présentent le plus de ressemblance spectrale avec [i]) et de toute syllabe initiale contenant [i] mais commençant par [t] ou [k] ⁽¹⁵⁾

Si l'on examine le spectre de [i] en commençant par les basses fréquences, on constate la différence de forme du premier formant au contact des trois occlusives sourdes. S'il possède sensiblement la même forme de 'U' renversé pour [i] suivant [p] et [k], il est précédé pour cette dernière consonne par une période non-voisée de 20 à 40 ms. Pour [i] suivant [t] la montée en fréquence semble plus lente avec une attaque de fréquence moins élevée. Pour le second formant, c'est le [i] précédé de [k] qui est le plus remarquable par l'atténuation sensible de ce formant et par son retard par rapport à la rupture de l'occlusion. A côté de ce type de réalisation, celles de la même voyelle au contact de [p-] et de [t-] paraissent nettes et facilement caractérisables. Ce même déséquilibre se retrouve autour de 3 kHz. Pour [i] suivant [t] apparaissent en outre de nombreuses résonances de 10 à 30 ms. à partir de 4 kHz qui semblent bien caractériser le [i] précédé de [t]. A cela vient s'ajouter le phénomène bien connu du retard de l'apparition de la voyelle à proprement parler après [t] et [k] auquel s'ajoute pour [ki] le retard de l'apparition des formants hauts par rapport à la portion voisée du premier formant. Donc, pour nous résumer, les réalisations de /i/ après les consonnes occlusives /p/, /t/ et /k/ diffèrent par la forme du premier formant, par l'intensité du second, par la présence ou l'absence de résonances dans les hautes fréquences et par la synchronie ou son absence entre l'apparition du premier formant et des formants supérieurs.

Pour ce qui est de la différenciation des voyelles [i],[e],[y] et [ø], et si l'on passe, comme nous l'avons fait, sur les différences relevées traditionnellement, il sera possible de constater le retard de l'apparition de F₂ des voyelles arrondies par rapport à [i] et [e] (ce qui ne serait pas le cas après [t]) ⁽¹⁶⁾. Les différences spectrales bien connues entre ces différentes voyelles suffiront à lever toute ambiguïté et dans tous les cas dans la mesure où les points de repère ne seront pas les valeurs moyennes traditionnelles, mais plutôt celles relevées dans le cadre syntagmatique.

5. CONCLUSION

Beaucoup d'études se sont contenté par tradition ou par doctrine de limiter l'analyse du signal de parole à une collection de valeurs moyennes centrales des unités sonores. Sans négliger l'apport des connaissances ainsi réunies, il semble évident qu'elles ne peuvent rendre compte des réalisations plus marginales. Les réunir reviendrait à accroître le nombre de ces données de références, les interférences au niveau de la reconnaissance sans se donner les moyens de pouvoir choisir entre elles. Ce n'est qu'au niveau de la chaîne sonore, dans laquelle ces unités s'insèrent, et en profitant du cadre syntagmatique où elles se trouvent que pourra être évitée la mise en parallèle d'une base de données où seraient privilégiées des unités standardisées comme illustrations d'unités de la forme, et la parole composée d'unités de substance fortement contextualisées.

(15) Enregistrement fait en chambre sourde, sur magnétophone Revox, analysés avec un "Digital Sonagraph 7800".

(16) Voir à ce propos, DURAND, P. 1985, pp.73-94