



HAL
open science

C'est "mm-hm, oui" ou "mm-hm, non"? Propositions pour une grammaire des composantes acoustiques des interactions nasalisées

Aurélie Chlébowski, Nicolas Ballier

► To cite this version:

Aurélie Chlébowski, Nicolas Ballier. C'est "mm-hm, oui" ou "mm-hm, non"? Propositions pour une grammaire des composantes acoustiques des interactions nasalisées. 6e conférence conjointe Journées d'Études sur la Parole (JEP, 31e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition), 2020, Nancy, France. pp.100-108. hal-02798522v1

HAL Id: hal-02798522

<https://hal.science/hal-02798522v1>

Submitted on 7 Jun 2020 (v1), last revised 23 Jun 2020 (v3)

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

C'est "*mm-hm, oui*" ou "*mm-hm, non*" ? Propositions pour une grammaire des composantes acoustiques des *interactions nasalisées*

Aurélie Chlébowski, Nicolas Ballier
Université de Paris, CLILLAC-ARP, F-75013 Paris, France
aurelie.chlebowski@hotmail.fr, nicolas.ballier@u-paris.fr

RÉSUMÉ

Cet article se propose d'envisager l'existence d'une grammaire spécifique aux *interactions nasalisées* (Chlébowski et Ballier, 2015). Notre proposition se fonde sur une annotation des composantes acoustiques de cette sous-catégorie de *sons non-lexicaux* (Ward, 2006) dans le corpus CID (Bertrand et al., 2008). Nous voudrions présenter les contraintes combinatoires et régularités qui semblent s'appliquer à ces composantes acoustiques, ainsi que discuter leur structuration. Les résultats préliminaires de l'analyse des composantes acoustiques semblent suggérer des plages de valeurs par défaut pour les réalisations des IN (notamment pour la durée). La violation de ces usages peut donner lieu à une analyse de type gricienne d'implicature.

ABSTRACT

A modest proposal for the pragmatic of nasal grunts in the CID corpus.

This paper proposes to consider that a specific grammar might govern the production of *nasal grunts* (Chlébowski et Ballier, 2015). This proposal is based on an annotation of the acoustic components of this sub-category of *non-lexical conversational sounds* (Ward, 2006) in the CID corpus (Bertrand et al., 2008). We would like to show the combinatory constraints and regularities that apply to these acoustic components as well as debate their potential organisation. Our preliminary observations suggest that certain ranges of realisations are expected (especially for duration), the violation thereof leading to Gricean implicature.

MOTS-CLÉS : interactions nasalisées, compositionnalité, CID.

KEYWORDS: nasal grunts, compositionality.

1 Introduction

D'abord exclus du champ de l'analyse linguistique (soit parce qu'ils étaient considérés comme événements marginaux, soit au prétexte qu'ils entraveraient la *bonne* production et compréhension du message principal), les sons « non-lexicaux » (entres autres dénominations, cf. Dingemans, 2020) ont, depuis, suscité la curiosité, de par leur omniprésence dans la conversation. Ces dernières décennies témoignent d'un regain d'intérêt significatif pour leur analyse, principalement motivé par le désir de rendre compte de leurs multiples rôles en interaction. On peut citer, entre autres, les études qui interviennent dans la compréhension ou le partage de l'état émotionnel, attitudinal ou cognitif du locuteur (Duez, 2001 ; Corley et Hartsuiker, 2011), celles qui s'intéressent au niveau de maîtrise d'une langue et à sa perception (Cenoz, 1998 ; Niebuhr et Fischer, 2019) ou encore celles qui traitent de la structuration des tours de parole (Peters et Wong, 2015). Quels que soient leurs

motivations, méthodes ou supports d'analyse (enregistrements, transcriptions, et même, corpus écrits), ces études se rejoignent pour ce qui est de leur approche fondamentalement onomasiologique : elles reposent sur des catégories majoritairement fonctionnelles, voire sémantiques (*backchannels*, *disfluences*, pour ne citer que celles-ci). Certaines ont proposé des hypothèses quant à la relation qu'entretiennent le sens et la forme des sons non-lexicaux (Duez, 2001 ; Clark et Tree, 2002 ; Dingemanse et al., 2013, entre autres) mais pas d'explication unificatrice sur la capacité qu'aurait une même forme sonore (telle qu'ils la posent *a priori*, hors toute considération de variation) d'endosser différentes fonctions ou de dénoter diverses significations (Ward, 2006). Néanmoins, ces études suggèrent que la production de ces sons n'est pas arbitraire : pour les employer correctement, il faut en apprendre et en comprendre l'usage. D'autres études se sont proposées de s'émanciper des catégories fonctionnelles susnommées. Compte tenu de la nature acoustique des *sons* non-lexicaux, Ward (2006) défend une approche sémasiologique, qui part des signes, et plus particulièrement d'une étude multidimensionnelle du signifiant sonore, pour remonter vers le concept. Il postule que ces sons sont régis par des règles différentes de celles qui s'appliquent au lexique : le sens transmis par un son non-lexical provient en grande partie de sa composition acoustique.

A défaut d'une sémantique, qui serait assez difficile à définir et sujette à caution pour la formulation de tests de perceptions qu'il faudrait établir (Chlébowski, 2016), nous proposons de discuter des régularités et contraintes qui nous paraissent à l'œuvre dans l'agencement des composantes acoustiques des sons non-lexicaux. Nous focalisons notre étude sur l'analyse des composantes des *interactions nasalisées* (ci-après, IN) dans le corpus CID (*Corpus of Interactional Data* ; Bertrand et al., 2008). La section suivante pose le cadre théorique, explicite les composantes acoustiques et problématise leur superposabilité. La section 3 présente le traitement des données tirées du CID. La section 4 esquisse la typologie des contraintes observées et la dernière section discute les résultats et conclut.

2 Cadre théorique

Cette section se propose de poser le cadre théorique dans lequel se situe notre étude, en explicitant d'abord la notion de *modèle compositionnel* (Ward, 2006), puis en proposant une représentation des composantes acoustiques des IN monosyllabiques telles qu'on peut les observer dans le CID.

2.1 Le modèle compositionnel de Ward (2006)

L'analyse acoustique des sons non-lexicaux n'est pas sans précédent et certaines études ont laissé entrevoir l'importance de la forme sonore de ces sons pour signifier une intention de sens. Néanmoins, la majorité d'entre elles se focalise sur les réalisations acoustiques d'une unique interprétation fonctionnelle de ces sons non-lexicaux. Les hypothèses formulées sont alors contraintes, selon nous, par les limites qu'impose cette catégorie unique en ce qu'elles ne permettent pas de rendre compte de ce qui est, ou n'est pas, partagé par différentes réalisations fonctionnelles (*backchannels*, *disfluences* ...) des sons non-lexicaux. Ward (2006) propose, pour l'anglais américain, de répertorier tous les sons non-lexicaux d'un corpus de taille moyenne, indépendamment des catégories fonctionnelles qui leur sont généralement prêtées. Son étude, principalement auditive, des formes acoustiques de ces sons est fondée sur une conception *décompositionnelle* de leur substance sonore. Il propose un modèle componentiel (*compositional model*) selon lequel les sons non-lexicaux sont décomposables en un nombre fini d'entités acoustiques (que nous appellerons ici des *composantes acoustiques*). Il relève une dizaine de

composantes (/h/ ou voix soufflée, voix craquée, nasalité, /o/, /a/, entre autres) qui se présenteraient de manière superposée (voire concaténée dans de rares cas) et dont le nombre de combinaisons est difficilement estimable. Selon ce modèle, chaque composante acoustique d'un son non-lexical porte une valeur sémantique qui s'exprime indépendamment du contexte dans lequel le son est produit. La signification du non-lexical, contrairement au lexical, n'est pas à chercher dans la succession des signifiants. La production des composantes acoustiques dans les sons non-lexicaux n'étant pas astreint à la linéarité du signifiant, il serait possible de dériver le sens d'un son non-lexical par l'inspection de sa composition acoustique. Bien que la saillance perceptive de certaines composantes soit notable (comme les variations de mélodie ou de durée), chaque composante d'un son non-lexical apporte bien une contribution au sens final du son non-lexical. Il n'est alors pas question de hiérarchiser les composantes mais plutôt de les traiter sur le même plan. Nous illustrerons l'intérêt du modèle compositionnel de Ward par l'exemple suivant :

- | | | | |
|------|--|------|--|
| (1a) | A : J'ai fait rentrer le chat.
B : Hein ?
A : J'ai fait rentrer le chat. | (1b) | A : J'ai fait rentrer le chat.
B : Ah ?
A : Oui, il miaulait à la porte. |
|------|--|------|--|

Posons que les composantes à l'œuvre dans les sons non-lexicaux dans (1a) et (1b) sont les mêmes, à l'exception de la qualité du segment prononcé par B (/ɛ̃/ vs. /a/). Sans entrer dans les détails, on remarque que la réponse de A est cohérente avec le son non-lexical produit par B. L'issue de la conversation varie considérablement selon le timbre vocalique choisi par B, ce qui semble accréditer l'idée qu'une seule composante acoustique de l'IN (telle qu'ici le timbre vocalique) puisse être signifiante. On perçoit ici les relations qu'entretiennent sens et forme sonore d'un son non-lexical mais aussi la reconnaissance de ces relations par le locuteur, qui est capable de les employer, et par l'interlocuteur, qui est capable de les décoder (Duez, 2001). Au-delà des applications sémantiques et pragmatiques impliquées par la sélection des composantes acoustiques dans la constitution d'un son non-lexical, nous estimons qu'il est d'abord nécessaire de déterminer la nature de ces composantes (et leur inventaire) dans une langue donnée. A la différence du modèle de Ward (2006), nous estimons que toutes les composantes acoustiques n'ont peut-être pas le même statut combinatoire.

2.2 Anatomie des composantes acoustiques des IN monosyllabiques autonomes

La représentation stratificationnelle proposée en Figure 1 (à gauche), qui superpose les composantes d'une IN monosyllabique *autonome* (voir section 3), nous paraît réaliser le programme de recherche de Ward. L'acte de production (ou de *vocalisation*) d'une IN contraint la présence de certaines composantes acoustiques, notamment : un son syllabique nasal, une durée, une valeur de fréquence fondamentale (F0), un degré d'intensité et un registre. Comme exposé dans Chlébowski (2020), nous considérons ces composantes comme *fondamentales* à la production d'une IN par rapport à d'autres qui sont, de fait, *optionnelles*. Chaque strate de cette représentation offre un éventail de valeurs possibles. Par exemple, la variation de la F0 peut être déclinée en contours *plat*, *montant* ou *descendant*. L'aspect catégoriel ou continu des valeurs que peut revêtir chaque composante n'est pas tranché (Ward, 2006). Il est tout à fait possible d'imaginer qu'il existe un continuum de *montant* à *descendant* (en passant par le niveau *plat*) au niveau de la 3^e strate, de *bouche fermée* à *bouche ouverte* pour la 1^{ère} strate (Chlébowski, 2020), ou encore continuum de *haut* à *bas* (en passant par le niveau *médian*) pour la 5^e strate.

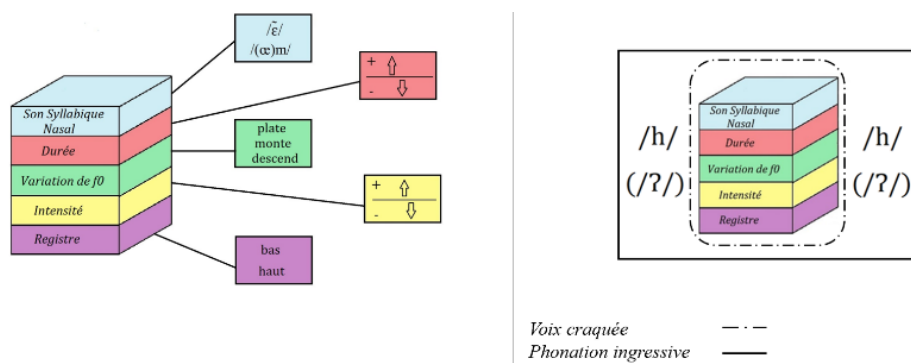


FIGURE 1 : Modèle stratificationnel des composantes *fondamentales* (à gauche) et modèle pour l'interaction avec les composantes *optionnelles* (à droite), pour les IN autonomes dans le CID.

Dans notre représentation, les composantes *optionnelles* (qui, selon nous, ne relèvent pas de paramètres pathologiques ou stylistiques) sont : la voix craquée, /h/ ou voix soufflée, la voix ingressive et le coup de glotte. Ces composantes sont sujettes à des contraintes physiologiques, voire phonotactiques : certaines seront davantage successives que simultanées. La propriété essentielle de leur combinatoire est la possibilité d'être potentiellement non-congruentes avec la durée totale des composantes principales de l'IN. Nous avons représenté ces composantes *additionnelles* à droite sur la Figure 1. La composante /h/, ou voix soufflée, peut apparaître en début ou en fin de IN. Lorsqu'elle est en position finale, elle est toujours discrétisable, contrairement à la position initiale où elle est susceptible de se propager sur les composantes principales de l'IN. Il semblerait que les coups de glotte aient une distribution comparable (sans la propagation). La phonation ingressive, quant à elle, commence avant les composantes principales de l'IN et se prolonge (fréquemment, après leur réalisation dans l'IN). Le craquement lui, est co-occurent avec le son syllabique mais présente la particularité de pouvoir être discontinu. Dans une optique compositionnelle, cette capacité de désynchronisation des composantes optionnelles permettrait de varier le sens transmis par une IN¹.

3 Constitution et traitement des données

Dans la continuité des travaux de Ward (2006), Chlébowski et Ballier (2020) ont proposé un protocole pour l'annotation des composantes acoustiques de 947 *interactions nasalisées (nasal grunts)*² sélectionnées dans le corpus CID. Les IN y sont définies comme une sous-catégorie de sons non-lexicaux qui partagent dans leur combinaison le trait acoustique de la *nasalité*. Les items

¹ Bien que cela complique les analyses, on pourrait proposer de considérer que ces composantes soient elles-mêmes divisibles en sous-composantes, par exemple : durée, intensité et position vis-à-vis des composantes principales.

² Chlébowski et Ballier (2015) emploient le terme *grunt* dans leur étude sur les *nasal grunts* en anglais Geordie, suivant (Ward, 2000), qui entérine à la fois le caractère non technique et stigmatisant du terme. La traduction de ce terme par « *grognement* » n'est pas, à notre connaissance, attestée dans cet emploi en français. Notons que l'ensemble de ses synonymes renvoie à des verbes de parole connotés négativement (*grommellement*, *bougonnement*, *marmonnement*, *ronchonnement*). Notre terme *interactions nasalisées* souligne l'aspect profondément interactionnel de ces sons.

susceptibles d’entrer dans cette catégorie dans le CID sont alors les suivants : *hein*, *han*, *hum* (ou *eh+mh*) et *mh* (ou *hm*). Ils soutiennent que l’observation d’indices acoustiques minimiserait la subjectivité qu’induit la perception auditive sur ces sons. Ils ont proposé un guide d’annotations de leurs composantes acoustiques, basées majoritairement sur des critères visuels, conduites sous Praat (Boersma et Weenink, 2019). Le tableau 1 ci-dessous récapitule la nature des observations proposées pour les composantes des IN. Les composantes *segment*, *registre* et *syllabation* ont, pour diverses raisons que nous ne re-exposeront pas ici, échappé à une analyse exclusivement visuelle. De plus, si le repérage de /h/ et de la phonation ingressive peut se faire visuellement, la distinction entre les deux nécessite l’écoute de l’annotateur. Il semble que seule la durée soit directement objectivable. Ils proposent une liste d’étiquettes restreintes, pour non seulement signifier la présence d’une composante mais aussi déterminer sa localisation dans l’IN. Ils soulignent que leurs annotations se veulent préliminaires à des études plus approfondies des corrélats acoustiques et de leurs interactions et reconnaissent que leur protocole d’annotation réalisé par un annotateur unique devra être soumis à validation inter-juges³.

D’autre part, Chlébowski et Ballier (2020) proposent d’annoter la position de l’IN dans l’énoncé du locuteur de l’IN (en début, milieu, fin d’énoncé ou comme étant *isolée* du reste de l’énoncé par des pauses) et en termes d’interaction entre locuteur et interlocuteur (voir la distinction en termes *self vs. others* dans Fruehwald (2016)). Nous avons retenu pour le présent article les IN prononcées à l’*isolée* dans l’énoncé du locuteur et qui peuvent donc être considérées comme *autonomes*⁴. En revanche, nous n’avons pas tenu compte de la position de l’IN en interaction.

Composantes	Type d’observation
Segment ; registre	Auditive
Syllabation	Visuelle et auditive
Distinction entre /h/ et phonation ingressive	
Variations de la F0 ; coup de glotte ; voix craquée ; /h/ ; phonation ingressive	Visuelle
Durée	Par mesure à partir des frontières des annotations
Intensité	

TABLE 1 : Degrés de la subjectivité des observations de Chlébowski et Ballier (2020)

4 Typologie des contraintes observées

Nous avons montré en 2.2 l’importance de contraintes physiologiques pour le statut superposable (craquement) ou concaténable (coup de glotte) des composantes acoustiques. Des

³ Les procédures envisageables sont, en fonction des composantes, visuelles, auditives et/ou automatiques. Pour la composante *syllabation*, par exemple, nous pourrions mettre en place des tests d’annotation croisée auditifs ou visuels (basés sur la recherche d’indices acoustiques). Une approche plus instrumentale pourrait être envisagée. Par exemple à l’aide de la détection automatique de pics syllabiques (De Jong et Wempe, 2009).

⁴ A l’inverse des IN prononcées en début, fin ou milieu d’énoncé (par exemple, *hein* en tant que marqueur de discours), les IN entourées par des pauses ne sont pas constituantes d’une unité prosodique plus importante qu’elles-mêmes.

contraintes d'ordre phonotactique peuvent également être observées. Ward (2006) et Chlébowski (2020) notent des combinaisons limitées pour certaines composantes. On retrouvera par exemple la succession /j/ + /e/ dans *yeah* (en anglais américain) ou la succession /œ/ + /m/ dans *hum* (en français dans le CID). A l'inverse, les séquences */e/+/j/ et */m/+/œ/, ne sont pas attestées et fort peu probables. Une partie de la distribution des formes des IN recensées pourrait s'expliquer par le jeu d'interrelations entre les composantes. Chlébowski et Ballier (2020) notent que la phonation ingressive se rencontre essentiellement avec les IN de type segmental /m/ et /ã/, mais pas avec celles de type segmental /ẽ/. Dans cette section, nous voulons montrer des tendances distributionnelles des composantes des IN mono- et dissyllabiques de type segmental /ẽ/ ou /m/ dans le CID.

4.1 Contraintes distributionnelles pour les monosyllabiques autonomes en voix modale

Nous concentrons nos premières observations sur la durée, la catégorisation du contour, et le sexe du locuteur pour les IN monosyllabiques autonomes en voix modale de type segmental /ẽ/ ou /m/. Les IN de type segmental /m/ semblent être préférées en position *isolée*, et ce, indépendamment du sexe du locuteur (tableaux de contingence en Figure 2). Le contour montant (M) est préféré pour les monosyllabes de type segmental /ẽ/ et les contours descendant (D) et plat (P) pour les monosyllabes de type segmental /m/ (p-valeurs pour le test exact de Fisher de 0,0009 et 0,002 pour les locuteurs et locutrices du CID, respectivement). En ce qui concerne les variations de durée (voir Figure 2, à gauche), nous rejoignons les résultats de Chlébowski et Ballier (2020) sur un autre jeu de données. L'hétéroscédasticité des données a été contrôlée avec un test de Levene (p-valeur < 0,001) sous R (R Core Team, 2017) et les résultats de l'ANOVA montrent une forte interaction de la durée avec le sexe du locuteur et le type segmental de l'IN. La catégorisation du contour ne semble pas jouer de rôle majeur dans les variations de la durée d'une IN. En moyenne, les locuteurs masculins produisent des IN plus longues que celles locutrices féminines et les IN de type segmental /m/ sont souvent plus longues que celles de type segmental /ẽ/. Nous remarquons aussi que les IN de type segmental /ẽ/ ne dépassent pas une certaine durée de production (jusque 250 ms.) alors que les IN de type segmental /m/ présentent une plus large plage de variation de durée (jusque 684 ms.). Certes, il est physiologiquement possible de réaliser un monosyllabe de type /ẽ/ d'une certaine durée, mais l'absence d'observations de durée supérieure à 250 ms. pourrait laisser penser qu'une réalisation de type /ẽ/ n'est pas compatible avec l'interprétation sémantique d'une durée plus longue. On voit par là qu'il pourrait exister des plages de valeurs par défaut qui sont susceptibles d'interagir avec l'interprétation.

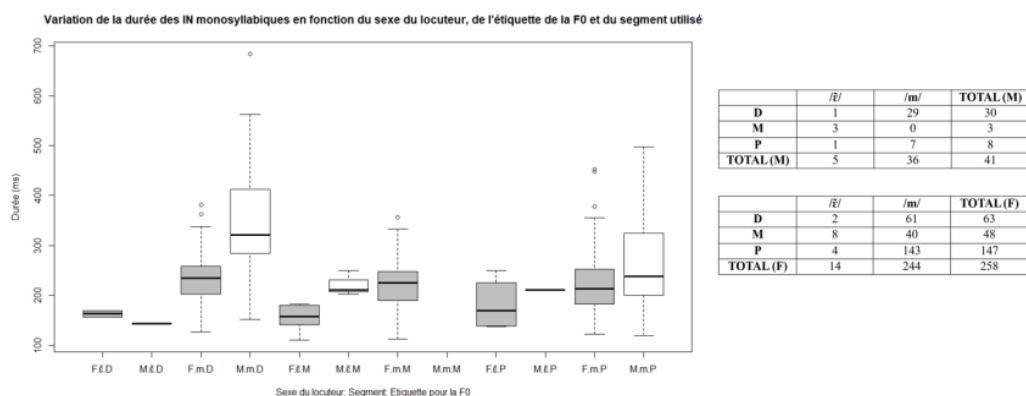


FIGURE 2 : Variation de la durée des IN monosyllabiques autonomes en voix modale en fonction du sexe du locuteur (F/M), du segment et de l'étiquette de la F0 (à gauche) et tableaux de contingence pour ces mêmes paramètres, à l'exception de la durée (à droite).

4.2 Dissymétrie de la durée dans les IN dissyllabiques

Nous ne reviendrons pas sur les difficultés du découpage syllabique exposées dans Chlébowski et Ballier (2020)⁵. Nous voudrions argumenter en faveur d'une attente des locuteurs en termes de réalisations des IN dissyllabiques en montrant que, par défaut, la durée de la première syllabe semble inférieure à celle de la seconde (Figure 3). Chlébowski (2020) propose que cette distribution s'explique non pas par des contraintes de production (il est physiologiquement possible de produire la première partie de l'IN plus longue que la seconde) mais plutôt par ce qui est de l'ordre des attentes pragmatiques. Allonger la première syllabe d'une IN serait interprété dans un contexte interactionnel de la même manière que pour les violations du principe de coopération de Grice (1975) : comme une implicature. S'il est conventionnel que la première syllabe soit plus courte, l'allonger produirait un effet non-conventionnel souvent sarcastique, glosable, entre autres, par « ah ouais, tu m'en diras tant ». En d'autres termes, cette interprétation sarcastique de l'IN fonctionnerait à la manière d'une méta-règle (Ballier, 2005) : si la durée de l'IN ne respecte pas la répartition attendue, l'interaction est d'ordre sarcastique (et pas dans l'approbation, pour reprendre la problématique du titre de notre article). Nous proposons de vérifier cette hypothèse par un exemple en contexte. Sur la Figure 3, les IN pour lesquelles le premier segment est nettement plus long que le second (environ 100 ms.) sont les IN 905 et 896.

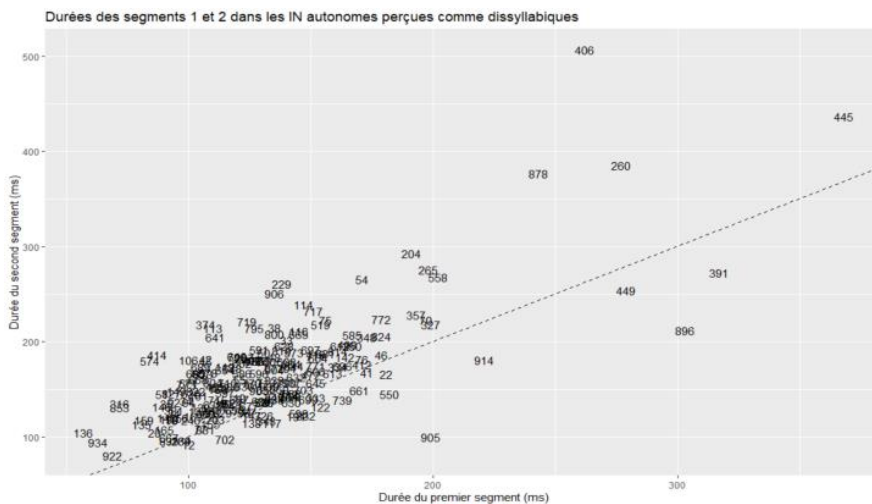


FIGURE 3 : Durées des premiers et seconds segments des IN autonomes perçues comme dissyllabiques (la nature de la réalisation de la découpe syllabique n'est pas prise en compte)

Nous focaliserons notre analyse sur l'IN 905 de la conversation entre les participants AP et LJ (bien que l'IN 896 soit passible d'une même analyse). Dans cette conversation, AP explique à LJ comment il s'est retrouvé à faire la cuisine, sans aucune expérience, dans un restaurant peu recommandable, à l'hygiène et aux méthodes douteuses. L'exemple (2) ci-dessous présente un extrait de cette conversation.

⁵ Nous soutenons néanmoins leur hypothèse qui pose que la présence des composantes /h/ ou *coup de glotte* en position médiane d'une IN fonctionne comme un marqueur de coupe syllabique (/m.hm/ et /mʔm/). Les résultats d'une première expérience d'annotation croisée (impliquant trois participants, étudiants en phonétique) semblent aller en ce sens, avec plus de 75% d'accord sur la forme dissyllabique des IN lorsqu'elles sont présentées avec un /h/ ou un coup de glotte médian. D'autres analyses sont en cours.

- (2) LJ (576,72) t'as fait la cuisine antillaise @@
 AP (577,90) ouais cuisine antillaise ben la cuisine antillaise en fait je croyais que
 c'était vachement dur quoi et en_fait j'étais un super cuistot
 LJ (582,75) @@
 AP (583,01) il suffisait d'ouvrir euh les les d- de piocher dans les trois congélateurs
 LJ (586,28) IN#905
 AP (586, 81) de prendre euh les les trucs tout faits

S'il n'est pas perceptible de prime abord dans la transcription donnée en (2), le sarcasme impliqué par cette IN l'est à l'oreille, selon nous. L'IN sarcastique « tu m'en diras tant, piocher dans les congélateurs, c'est pas être un super cuistot » semble alors tout à fait à propos pour ponctuer le côté absurde du contexte.

5 Discussion et conclusion

Cet article, s'il ne constitue que des propositions et observations, a néanmoins le mérite de proposer une interprétation systémique des résultats obtenus. Nous avons voulu montrer le caractère superposable et/ou concaténable des composantes acoustiques à l'œuvre dans les IN du CID, en rendant compte de certaines contraintes et régularités.

Au-delà de l'importance des contraintes physiologiques qui semblent diviser les composantes en ce qu'elles seraient *fondamentales* ou *optionnelles*, nous avons suggéré l'importance que joue la sémantique dans leur combinatoire. Il est nécessaire de distinguer les impossibilités physiologiques de certaines productions (à la manière des articulations jugées impossibles des cases noires du tableau de l'API) et leur improbabilité, qu'elle soit agrammaticale (cas de *mh* + *euh*) ou apragmatique (produire la première syllabe d'une IN dissyllabique plus longue que la seconde). Les composantes acoustiques des IN seraient sujettes à une sorte de « grammaire » (Chlébowski, 2020) où la sélection des composantes serait principalement motivée par des choix sémantiques susceptibles d'exploitations pragmatiques, comme on vient de le voir. Notre analyse s'inscrit dans un courant plus général qui milite en faveur d'une grammaire de l'interaction (Ginzburg et Poesio, 2016).

Nous avons bien conscience qu'un certain nombre de questions restent en suspens. Entre autres, pour ces IN, selon les langues considérées, se pose toute la question de la grammaticalisation des composantes acoustiques (l'utilisation et l'investissement sémantique de telle ou telle composante acoustique). Les contraintes de production (ou les habitudes articulatoires ?) ne sont pas nécessairement les mêmes en français et en anglais, ou même entre variétés d'une même langue (Chlébowski, 2020). Le CID étant un corpus de français méridional, il ne semble pas exclu d'avoir des réalisations septentrionales qui soient différentes. Est-ce à dire pour autant qu'il n'y pas d'universel pour ces *interactions nasalisées* ? Nous défendons que seule une approche intégrative (Ward, 2006) des études sur les sons non-lexicaux pourrait nous permettre de mieux comprendre leur *modus operandi*. En ce sens, nous ne considérons pas notre approche comme *meilleure* mais comme *complémentaire* de toute autre étude sur le sujet.

Remerciements

Nous tenons à remercier Roxane Bertrand pour nous avoir donné accès au CID et les deux relecteurs anonymes pour leurs commentaires.

Références

- BALLIER, N. (2005). De l'exception à la régulation des exceptions pour la phonologie de l'anglais: d'un système de métarègles?. *Faits de Langues*, 25(1), 245-253.
- BERTRAND, R., BLACHE, P., ESPESSER, R., FERRE, G., MEUNIER, C., PRIEGO-VALVERDE, B., & RAUZY, S. (2008). Le CID-Corpus of Interactional Data-Annotation et exploitation multimodale de parole conversationnelle. *Traitement Automatique Des Langues*, 49(3), 105-134.
- BOERSMA, P. & WEENINK, D. (2019). *Praat: doing phonetics by computer*. Version 6.1.06.
- CENOZ, J. (1998). Pauses and Communication Strategies in Second Language Speech. *Reports-Research*, 143, 1-11
- CHLÉBOWSKI, A. (2016). The Meaning of “Nasal Grunts” in the Necte Corpus. A Preliminary Perceptual Investigation. *Research in Language*, 14(1), 43–59.
- CHLÉBOWSKI, A. (2020). A Semasiological Approach to Non-Lexical Conversational Sounds: Issues, Benefits and Impact. In Proc. *Laughter and Other Non-Verbal Vocalisations Workshop*
- CHLÉBOWSKI, A., & BALLIER, N. (2015). “Nasal grunts” in the NECTE corpus, Meaningful interactional sounds. In Proc. *EPIP4-4th International Conference on English Pronunciation: Issues & Practices*, 54–58.
- CHLÉBOWSKI, A., & BALLIER, N. (2020). A Manually Annotated Resource for the Investigation of Nasal Grunts. In Proc. *LREC 12th*
- CLARK, H. H., & TREE, J. E. F. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, 84(1), 73–111.
- CORLEY, M., & HARTSUIKER, R. J. (2011). Why *um* helps auditory word recognition: The temporal delay hypothesis. *PloS One*, 6(5), e19792.
- DE JONG, N. H., & WEMPE, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, 41(2), 385–390.
- DINGEMANSE, M. (2020). Between sound and speech: Liminal signs in interaction. *Research on Language and Social Interaction*, 53(1), 188–196.
- DINGEMANSE, M., TORREIRA, F., & ENFIELD, N. J. (2013). Is “*Huh?*” a universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PloS One*, 8(11), e78273.
- DUEZ, D. (2001). Signification des hésitations dans la parole spontanée. *Revue Parole*, 17–18.
- FRUEHWALD, J. (2016). Filled pause choice as a sociolinguistic variable. *University of Pennsylvania Working Papers in Linguistics*, 22(2), 41–49.
- GINZBURG, J., & POESIO, M. (2016). Grammar is a system that characterizes talk in interaction. *Frontiers in psychology*, 7, 1938.
- GRICE, H. P. (1975). Logic and conversation, syntax and semantics. *Speech Acts*, 3, 41–58.
- NIEBUHR, O., & FISCHER, K. (2019). Do not hesitate!—Unless you do it shortly or nasally: How the phonetics of filled pauses determine their subjective frequency and perceived speaker performance. In Proc. *Interspeech 2019*, 544–548.
- PETERS, P., & WONG, D. (2015). Turn management and backchannels. In *Corpus pragmatics: A handbook* (408–429). Cambridge University Press.
- R CORE TEAM (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- WARD, N. G. (2000). Issues in the transcription of English conversational grunts. In Proc. *1st SIGdial Workshop on Discourse and Dialogue*, 29–35.
- WARD, N. G. (2006). Non-lexical conversational sounds in American English. *Pragmatics & Cognition*, 14(1), 129–182.