A Semasiological Approach to Non-Lexical Conversational Sounds: Issues, Benefits and Impact

Aurélie Chlébowski
Université de Paris
CLILLAC-ARP, F-75013 Paris, France
aurelie.chlebowski@hotmail.fr

Abstract

This paper proposes to consider a semasiological approach to *non-verbal vocalisations*. We claim that an acoustic analysis of the components of these sounds is needed to complement the findings of earlier studies. We propose that part of the information conveyed by these sounds comes from their acoustic components and that these components might be subjected to what resembles grammatical rules. Semantic issues are discussed at the end of the paper.

1 Introduction

The last decades testify to a renewed interest for the understanding of human *non-verbal* communication. As regards the analysis of *non-verbal vocalisations*, most of the studies, all with their own research question (Tottie, 2019), followed an onomasiological approach: they focused on the concept first, *e.g.* disfluencies, filled pauses, backchannels… Still, how these sounds convey meaning remains a complex issue. Nigel Ward (2006) proposed to depart from the standard "functional" categories and performed a semasiological analysis on what he called *non-lexical conversational sounds* (hereafter N-LC sounds). He claimed that the meaning conveyed by N-LC sounds follows a *compositional model*. In this model, each acoustic component of an N-LC sound (*e.g.* duration, segment nature, voice quality…) bears a meaning. As components combine to create an N-LC sound, so does the number of meanings. The final meaning of an N-LC sound could therefore be inferred from the combination of its components.

This paper offers to complement the findings of earlier studies with an analysis of the acoustic components of *nasal grunts* (hereafter NG; Chlébowski and Ballier, 2015); a sub-category of N-LC sounds. The remainder of this paper is as follows: section 2 summarizes our procedure for- and issues of- the acoustic analysis of NG. Section 3 presents our findings: NG are not as spontaneous as one might think, but rather, they seem to follow "grammatical rules". Section 4 discusses open-ended questions as regards semantics. Section 5 recaps issues remaining to be addressed and concludes.

2 Annotations

This section outlines our procedure for the annotation of the acoustic components of NG and that of the distribution of NG in conversation[1].

2.1 Corpora

We focused our analysis on NG from three corpora. 947 NG were investigated in the French Corpus of Interactional Data (hereafter, CID; (Bertrand et al. 2008). 198 NG and 332 NG in selected files from the Santa Barbara Corpus of Spoken American English (hereafter, SBC; Du Bois et al., 2000) and from the Newcastle Electronic Corpus of Tyneside English corpus (hereafter NECTE; Corrigan et al., 2001), a corpus of Geordie English. According to Chlébowski and Ballier (2015), *nasal grunts* are "words which have no "clear meaning" (Ward, 2000: 29) but possess a nasal feature" (p.54). Orthographic tokens that fit into the NG category in our corpora are: *hein, han, hum, ehm, mh, mmhm, hm* and the like.

---

[1] Annotations and acoustic analysis were made with Praat software (Boersma and Weenink, 2019). Explanations of the acoustic analysis and annotation guidelines are detailed elsewhere (Chlébowski and Ballier, 2020).

2.2 Annotation of the Acoustic Components

We designed annotation guidelines that do not only code the presence or absence of a set of acoustic components but also their positions as regards each other. Our annotation procedure was meant to reduce the biases that could be induced by auditory perception as much as possible. We endeavoured to provide a restricted number of labels based on visual cues that could be quickly identified. Segments, syllabification, variations of the fundamental frequency (f0), register, voice quality (*i.e.* creaky voice and ingressive phonation), /h/ and medial glottal stop were investigated[2].

2.3 Annotation of the Distribution of NG in Conversation

The annotation of the distribution of NG in conversation was two-fold.

First, we annotated the position of the NG in the speaker's own speech. Results show that NG can appear not only at the beginning of, end of, or inside an utterance but also between utterances. In the latter case, the NG can be considered as autonomous since it is not construed within wider intonational units (*i.e.* it is surrounded by silences). The distribution of NG as regards interactions was coded in terms of self *vs.* other occurrences (Fruehwald, 2016).

2.4 Issues we Faced

The analysis of the nature of the segment, register and syllabification was complex.

Distinguishing a nasal vowel (*e.g.* /ɛ̃/) from a nasal consonant (*e.g.* /m/) on the spectrogram was not as easy as expected. Surprisingly, resorting to auditory perception was not of any help in some cases. We assume that this might come from the fact that nasal vowels and consonants in NG are not distinct phonemes but rather constitutes a *continuum*[3] from "closed mouth" to "open mouth" nasal sounds [4] .Therefore, when the mouth is halfway between the two, it might be difficult to determine the nature of the segment relying on auditory perception only, *i.e.* without video data[5].

Register used by the speaker (either inside or outside his/her comfort zone) cannot be inferred by simple visual cues, and our annotator relied on auditory perception for this component.

The complex issue of syllabification is not solved with NG. The insertion of either /h/ or a glottal stop between two segments (*i.e. openness vs. complete closure* of the vocal folds) seems to be a strong cue for determining syllable boundaries (*e.g.* /m.hm/ *vs.* /mʔm/). However, NG can still be perceived as disyllabic even though they lack such components in medial position[6] and in such cases, the visual inspection of acoustic cues for syllable count may conflict with auditory perception.

3 A Grammar of NG?

Our preliminary results suggest that the acoustic components of NG follow grammatical rules[7].

Some components are *essential* to the production of an NG, namely: a (somehow) *nasal* segment, a duration, an amplitude, a f0 value and a register. These components are simultaneous and superimposed (Ward, 2006) and can be represented in a stratificational model in which each stratum can offer a range of possible values (Figure 1,).
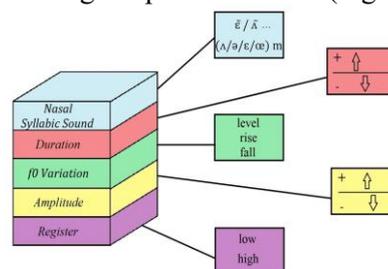

Figure 1: Components essential to the production of monosyllabic autonomous NG in our corpora

Other components would be *additional* in the sense that they might be used to diversify the content of the information conveyed. In this case, the position of the acoustic components is essential. Glottal stops and /h/ can appear at the beginning or end of monosyllabic grunts, while ingressive phonation would begin before the NG and spread over it. Creaky voice occurs on the NG and is mobile *i.e.* instances of creaks can come and go on the NG.

---

[2] Given our compositional model, duration can be construed from any aforementioned component.

[3] A recurring idea in Nigel Ward's work.

[4] Also suggested in Dingemanse et al. (2013) with the idea of "closed-mouth variants" for *huh?* sounds.

[5] This kind of analysis might be conducted on the CID for it is designed to investigate multimodality.

[6] This specificity was already noted by Ward (1998) on Japanese backchannels and was characterised as a "strong vibrato" (p.466).

[7] It should be borne in mind that our analysis is preliminary to deeper acoustic analysis. We also acknowledge the need for validation procedures of our results as well as more robust cues for detection of some components, *e.g.* syllabification.

Second, segments of different nature can combine in a fixed manner. The vowels /ʌ /or /ə/ in American English, /ɛ/ (sometimes / ə/) in Geordie English and /œ/ in French always appear before the consonant /m/.

Third, although the syllable count was complicated, we explained in section 2.4 that either /h/ or a glottal stop can be used as indicators for syllable boundaries. In cases where those additional components were missing but the NG was still perceived as disyllabic, the first perceived syllable seems always shorter than the second one. Further analysis is underway to determine the interactions at the level of: 1) fundamental components (*e.g.* can a nasal vowel segment be of a certain duration?), 2) additional components (*e.g.* can creaky voice and ingressive phonation combine?) and 3) the interface between fundamental and additional components (*e.g.* can /h/ appear at the beginning of /ʌ, ə, ɛ, œ + m/ NG?).

4   Open-Ended Questions

This section discusses the potential impact of our findings. We allude to the distinctive functions of the acoustic components of NG and the semantic implications. We also offer a parallel with laughter.

4.1   Distinctive Components *vs.* Production Constraints

Our analysis of the acoustic components of NG was meant to acknowledge as many acoustic components as possible with simplified labels. Nonetheless, it is worth considering the sound inventory of a given language before trying to assign semantic values to these components.

For instance, glottal stops are often realised at the onset of vowel initial N-LC sounds in English (as evidenced in Luthy, 1983). Therefore, the production of a glottal stop in onset position of our vowel initial NG from NECTE and SBC might be the outcome of an articulatory constraint and not a component used for semantic purposes.

Conversely, French is a language that does not seem to presuppose the need for glottal stop as onset of vowel initial words. A glottal stop occurring as onset of vowel initial NG might be used for communicational purposes (Malécot, 1975). This suggests that some additional components may be grammaticalized differently across languages.

4.2   Semantic Implications and Implicature

Ward (2006) and Chlébowski and Ballier (2015) made suggestions on the semantic values conveyed by the acoustic components of NG.

Our preliminary results suggest the possibility of consistent interactions between components that might confirm previous semantic hypotheses. For instance, the fact that creaks could be observed across an NG might be an argument in favour of the theory that they are used when the speaker needs to withdraw from the subject under discussion (Ward, 2006; Chlébowski and Ballier, 2015) and might be related to planning phases (*i.e.* multiple occurrences of creaks in a given NG would denote how many times the speaker needed to withdraw from conversation to formulate his/her thought). In that sense, /h/ or a glottal stop found in onset position of an NG might convey a different meaning than when found in coda position.

In everyday interactions, NG are usually processed unconsciously. Nevertheless, it does happen that an NG draws attention by implicature[8]. Given our grammatical perspective, the combination of some acoustic components of NG might not fit the *norms* and results into irony or sarcasm, *e.g.* making the first syllable longer in a disyllabic NG. There might also be linguistic, social, and cultural constraints as to the *canonical* or *expected* ways to produce an NG, *e.g.* contexts in which an NG should be produced to conform to the cooperative principle (Grice, 1975). Some NG might sound irrelevant, inappropriate, and even rude in specific contexts.

4.3   Laughter: Another Component of NG?

Our annotator noticed that some occurrences of laughter with the closed mouth (/m/ laughter) were mistaken for *mh* NG by the annotators of the CID. We did not investigate this specificity yet, but our annotator noticed some acoustic differences between NG and occurrences of laughter. Vocalisations involving multiple segments (*e.g.* /m.m.m.m.m/) along with insertion of /h/ and/or glottal stops between the segments and/or alternance between /h/ and ingressive phonation might suggest occurrences of laughter. Elevation of f0 (related to *register* in our study) reported by, for instance, Makagon et al. (2008) in laughter does not seem to be robust argument as regard our own study. NG can be uttered in high register without sounding *laughter-like*.

---

[8] See for instance Clark and Tree (2002) on *um* and *uh*.

There is a thin line between NG and laughter. The acoustic components involved in laughter seem to be highly comparable to that involved in the production of NG (Ward, 2006; Dodane et al., 2012). We hypothesize that there is no frontier between laughter and NG and even N-LC sounds. Given our compositional model, we believe that any occurrence of N-LC sound is the aggregate of its acoustic components. N-LC sounds are not only characterised by their segment quality (*e.g.* [a] *vs.* [o] *vs.* [m]) but also by every acoustic component they are composed of (*e.g.* creaky voice, /h/, glottal stop…), along with their combinational properties and variability. In that respect, NG and laughter should not be considered as different occurrences of N-LC sounds but rather as different clusters. Therefore, what we call *laughter* might either come from a certain aggregation of the acoustic components, or, *laughter* might be another component itself (like creaky voice) which would combine with other components (such as [a], [o], [m], glottal stop and ingressive phonation).

5   Discussion and Conclusion

We proposed to consider a semasiological approach to NG. Our preliminary findings suggest that the acoustic components of NG follow a specific grammar. Our semantic hypotheses can be formulated as follows: 1) distinctive functions of the components depends on the language under scrutiny, 2) the components of NG convey meaning and trigger implicature, 3) NG and lagher might be aggregates of the same acoustic components. Our work is still in process, but we believe, as Dodane et al. (2012) suggested for laughter, that acoustic components of N-LC sounds mostly originate from physiological characteristics that are used for communication purposes and are governed by linguistic processes.

Acknowledgments

I wish to thank Nicolas Ballier, as well as the reviewers, for their valuable comments on an earlier version of this paper

References

Bertrand, R., Blache, P., Espesser, R., Ferré, 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 [Computer program].* Version 6.1.06.

Chlébowski, A. and Ballier, N. (2015). Nasal grunts" in the NECTE corpus, Meaningful interactional sounds. In *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 *LREC 12th*

Clark, H. H. and Tree, J. E. F. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, 84(1), 7-111.

Corrigan, K., Allen, W., Beal, J., Maguire, W., Moisl, H., & Rowe, C. (2001). *Newcastle Electronic Corpus of Tyneside English Corpus.*

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.

Dodane, C., Sauvage, J., Hirsch, F., & Barkat-Defradas, M. (2012). Is children's laughter related to their language development. In 3*rd   Interdisciplinary Workshop on Laughter and other Non-Verbal Vocalisations in Speech*, 2-3.

Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., & Martey, N. (2000). *Santa Barbara Corpus of Spoken American English. CD-ROM*. Philadelphia: Linguistic Data Consortium.

Fruehwald, J. (2016). Filled pause choice as a sociolinguistic variable. *University of Pennsylvania Working Papers in Linguistics*, 22(2), 41-49.

Grice, H. P. (1975). Logic and conversation, syntax and semantics. *Speech Acts*, *3*, 41-58.

Luthy, M. J. (1983). Nonnative Speakers' Perceptions of English "Nonlexical" Intonation Signals. *Language Learning*, 33(1), 19-36.

Makagon, M.M., Funayama, E. S., & Owren, M. J. (2008). An acoustic analysis of laughter produced by congenitally deaf and normally hearing college students. *The Journal of the Acoustical Society of America,* 124(1), 472-483.

Malécot, A. (1975). The glottal stop in French. *Phonetica*, 31(1), 51-63.

Tottie, G. (2019). From pause to word: *Uh, um* and *er* in written American English. *English Language & Linguistics*, 23(1), 105-130

Ward, N.G. (1998). The relationship between sound and meaning in Japanese back-channel grunts. In *Proceedings of the 4th Annual Meeting of the (Japanese) Association for Natural Language Processing*, 464-467.

Ward, N.G. (2000). Issues in the transcription of English conversational grunts. In *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.