CHAPTER 7

Functional motivations for the sound patterns of English non-lexical interjections

Gina Joue
Institute of Neuroscience, Trinity College Dublin

Nikolinka Collier
Department of Computer Science, University College Dublin

Non-lexical interjections have been the focus of much research, but their apparent complexities, functional variations, and lack of content have led to different approaches to annotation and classification. This paper argues that they are discourse particles that function with strong cognitive linguistic bases and regularities in communication. They may have appeared as so complexly varied because they were assumed to be paralinguistic phenomena. We investigate our claims on two spontaneous speech corpora of English. In classifying these interjections into a taxonomy of discourse functions and using methods based on Phonology as Human Behavior, we find an interaction between the sound pattern of an interjection and its function in discourse, supporting our claim that non-lexical interjections are important linguistic phenomena.

1. Introduction

Our speech is replete with vocalizations like *mm-hm*, *and so* or short phrases like *you know* in contexts where their function or significance is disputed. Classic linguistic approaches note that these interjections do not easily fit into a single category but fall under different part-of-speech categories (e.g. adverbs, connectives, or short sentences). Work in traditional frameworks has tended to pigeonhole interjections as paralinguistic or noisy phenomena (Quirk et al. 1972) and ignore intuitions to the contrary that might indicate that these units share similar roles or functions.

An approach to language emphasizing its dynamic interactions (not rule-governed formalisms) brings us to the realization that interjections may function as signals exchanged in the process of negotiated understanding that is communication.

2. Non-lexical interjections in contemporary linguistics

Interaction strategies have often been key to understanding the contribution of paralinguistic signals (e.g. laughter and coughing) to communication. Some experts are convinced that interjections are yet another instance of a paralinguistic event (Quirk et al. 1972; Goffman 1981). Others argue for the linguistic status of interjections but have treated them primarily as aids in segmenting utterances (speech counterparts of punctuations) and in modelling the conversation flow (Hirschberg and Litman 1993; Shriberg 1994; Heeman and Byron 1998; Heeman et al. 1998; Lickley 2001).

Research in the last two decades has suggested that these units are linguistically relevant phenomena. However, linguists of different specializations and research expertise – computational (Heeman et al. 1998; Traum 1994), cognitive (Fischer 2000; Clark and Fox Tree 2002; Fox Tree and Schrock 1999), sociolinguistic (Schegloff 2001), applied (Fischer 2000; Lickley 2001), and theoretical (Schiffrin 1987; Fraser 1999) – remain in disagreement about their classification both in regard to identifying what belongs in their class and in terms of finding language to describe or designate their class.

Thus, finding appropriate terminology for referring to these units has been a crux of many previous considerations. What we have called *non-lexical interjections* have also been classified under *discourse markers*, *discourse particles*, and *cues*. However, for example, no consensus exists on what items or phenomena should be included in the class *discourse particles*. Nonetheless, even advocates of differing terminologies agree that they are studying linguistic phenomena: language devices which are generally syntactically detached and help elucidate the discourse structure.

With further complications, non-lexical interjections seem *lexicalized* in the sense that they have accepted denotations. The non-lexical status of these items may result from the variation they seem licensed to undergo; a license not as freely given to more commonly accepted lexical items. *Mm-hm*, *mmm-hmmmm*, *mmmm,* and *uh-huh*, all seem to function much like (the indisputably lexical) *yes*. Whether or not all of these variants are really the same, these categorization difficulties and the lack of a predictable structure for interjections contribute to arguments for their non-lexical and paralinguistic status.

We will argue that non-lexical interjections are more regular than previously thought: their variants can be simplified to a base form. Any study of interjections must take into account how they can function as a part of discourse. Moreover, consistently ignoring the discourse function of these terms itself creates and exaggerates the impression that their sound makeup is unstructured.

Our analysis supporting the functional character of non-lexical interjections will draw on principles from Phonology as Human Behavior (PHB) (Diver 1979; Davis 1984 [1987]; Diver 1995; Tobin 1997). Our recognition of the importance of communication and human perceptual and production constraints on language accords well with the "human factor" stressed in PHB and Columbia School Linguistics generally.

3. The phonology of non-lexical interjections

Little research has been done on the phonology of non-lexical interjections. Most of the research has been informal observations or speculative commentary. The more focused studies have been based, predominantly, on sound symbolism. For example, researchers have casually but consistently noted that non-lexical interjections commonly involve infrequent or illegal phonotactic combinations (Abelin 1999; Montes 1999; Ward 2000). They have also noticed that interjections in different languages share phonetic similarities.

In examining Icelandic, English, Polish, Hungarian, Finnish, Ososo, Malagasi, and Slovenian interjections, Abelin (1999) found that non-lexical interjections involve mostly labial or alveolar sounds. She also argued that their sounds have intrinsic (iconic) meanings. Like Abelin, Ward (2000) appeals to sound symbolism. In studies of American English (Ward 2000) and Japanese (Ward 1998), he proposes that the sound structure of non-lexical interjections (or "grunts", as he terms them) reflects the attitude or knowledge of the speaker. In a more detailed phonetic analysis of American English interjections, Ward (2006) builds further on his argument that the meanings of grunts may be predicted given their phonetic characteristics. He speculates that these phonetic characteristics perhaps emerge from cognitive factors. However, he proposes very specific sound symbolic mappings and does not focus on the motivations of the relationships among form, function, and meaning.

Rather than submitting to sound symbolism, we argue that the phonotactic skewings have clear cognitive bases. In the style of PHB, we can identify functional motivations for the sound structure of non-lexical interjections in terms of the "grounding process" (a term coined by Clark and Schaefer, 1989), where "grounding" is the negotiation of understanding among conversation participants. In so doing, we will gain insight into the role of non-lexical interjections in discourse.

4. Interjection sound patterns: The hypotheses

PHB posits that the non-random sounds in languages emerge from the (sometimes) conflicting goals of minimizing production efforts while negotiating efficient communication (cf. Tobin, this volume). We can reason then that less crucial changes in the discourse interaction will involve non-lexical interjections with less *marked* sounds and sound patterns. For example, we can explain why most pause fillers (the type of non-lexical interjections used to keep other participants from interrupting) seem to consist primarily of vowels (e.g. *uh*, *ah*, *eh*, *oh*). Vowels require less effort to articulate than consonants. The pause fillers are intermediary moments between heavier, information-laden articulatory gestures; it seems natural for the speaker to expend only minimal effort in their production, just enough to hold the floor (or not).

A relative scale of articulatory difficulty and perceptual saliency for particular sound patterns would, we think, indicate the degree to which a speaker desires to hold the floor. PHB and biomechanical studies (see Kim 1995 for a brief review) agree that the apex of the tongue and the lips are among the most flexible oral articulators. This explains Abelin's (1999) observation that interjections generally tend to involve sounds produced by either the lips or the apex of the tongue.

More formally, we can consider the discourse functions of non-lexical interjections to be either *static* or *dynamic*. Those that are *static* do not change the current beliefs or understanding of the participants or the intentional direction of the topic flow, but indicate, for example, the speaker's attendance to the conversation. We hypothesize that these static-functioning non-lexical interjections will overall be much simpler and vary less phonetically than interjections indicating more *dynamic* participation, which unlike static interactions, updates the beliefs and knowledge of the participants. In other words, static-function interjections will tend to involve the less salient sounds and/or more easily articulated sounds (what will be referred to as *unmarked* sounds). This implies a comparatively limited phonetic inventory and very simple syllable structures (most likely monosyllables). Analogously, we hypothesize that dynamic-functioning interjections indicate a speaker's willingness to increase articulatory effort for greater communicative holds and/or to produce particles with greater perceptual distinctions (or marked sounds).

5. DiSPEL (DIScourse Particle Expert modeL): A functional taxonomy

In order to test these hypotheses, we created a functional taxonomy for interjections that is simple enough for computational purposes but which also sufficiently captures the functions of interjections as discourse particles.

As discourse particles, non-lexical interjections constrain the discourse process. They represent the speaking intentions of participants, making utterances more predictable and thus understandable. A trend in discourse particle models has been to separate the discourse-managing processes (the process of grounding) from the flow of topic information. Within our DiSPEL framework all discourse particles, including non-lexical interjections, function as grounding units. Our taxonomy, developed from one presented by Traum (1994), also treats the grounding process as the set of possible contributions constrained by the information and intention of the participants.

Four factors in DiSPEL describe the management of communication or discourse moves, as shown in Figure 1.

Discourse Moves	Factors			
	Topic	Relation	Projection	Evaluation
Acknowledgement	old	initiating related	other	negative positive neutral
Expansion	new	related	self, other	
Correction	old	initiating related	self, other	negative

Figure 1. DiSPEL discourse management moves

1. **Topic** indicates whether the utterance is related to a "new" or a previously spoken-of ("old") topic.
2. **Relation** is the hierarchical interdependency among the utterances in a dialogue. Dialogues include main topics and subtopics each involving dependencies among the utterances. Utterances relate to each other not only by topic but also as the result of the communication strategies speakers use. For example, questions expect answers; a greeting is typically returned. Answers and return greetings are "related" to previous utterances. After greetings are exchanged, one of the participants would often "initiate" a new topic.
3. **Projection** indicates whether a speaker is making a statement based on his own knowledge ("self") or based on something the "other" participant has said.
4. **Evaluation** is the type of assessment a speaker gives: that is, "positive" (affirmation), "negative" (disagreement), or "neutral" (registering attention).

These four factors delimit the moves that discourse particles realize. According to grounding theory, the moves monitor and manage the discourse process. We have

classed the examples in our study based on three discourse moves: (1) *acknowledgment*, (2) *expansion*, and (3) *correction* (negative assessment).

Acknowledgment indicates the assessment of information provided by the other speaker (thus projection is always "other"). The type of evaluation an acknowledgment move has (positive, negative or neutral) dictates the types of utterances that are expected to follow. For example, *uh-huh* (from the TRAINS 91 corpus, Allen & Schubert 1991), constitutes a positive acknowledgment move as it indicates that speaker M agrees with speaker S's utterance. Cases in which *uh-huh* is followed by a negation like *but* are also classified as acknowledgement (even if partial) because the negation indicates a change in turn.

(1) S: that only takes a couple of minutes
M: *uh-huh*

Negative acknowledgment describes disagreement. *Uhh* in this TRAINS 91 example indicates dissent:

(2) M: if there are oranges at the orange juice factory
S: *uhh* no

Neutral acknowledgment describes attentiveness (assessment without agreement or disagreement). In Example (3), also from TRAINS 91, speaker S is suggesting a plan to participant M, who utters *m-hm* to encourage speaker S to continue rather than to actively agree or disagree with the plan.

(3) S: with this plan we can get the bananas to Corning at 1
M: *m-hm*
S: and we have the OJ in Corning at 6
M: okay

Expansion moves (by a speaker or an interlocutor) extend a previously spoken topic with new information. An example is provided in (4):

(4) M: so *um* do I tell you what to do at this point

Lastly, *Correction* is a move that partially or entirely changes an utterance or topic that has been discussed. It is always a negative type of evaluation. In the following example, both *oh* and *whoops* function as self-corrections:

(5) M: keep the engine there if it's not needed anywhere else send
S: well we only have oranges in Corning
M: *oh* I'm sorry right *whoops*

Oh and *whoops* signal speaker M's realization of having overlooked part of the situation when giving an initial plan.

For our tests, we classed interjections first according to the discourse move taxonomy and then distinguished between those that function statically or dynamically in the discourse. We tested our hypotheses in two dialects of English.

6. The corpora: Choosing and post-coding

Finding appropriately annotated public-domain corpora with consistently annotated non-lexical interjections was difficult because early research in spoken language systems filtered out interjections as irrelevant to the process of communication. This problem restricted our corpus choice to the TRAINS 91 dialogues (Allen and Schubert 1991) and the HCRC Map Task Corpus (Boyle 1990, Carletta et al. 1996).

The TRAINS 91 corpus is a collection of 16 task-oriented Wizard of Oz game dialogues among gender-balanced and regional-dialect-balanced American English speakers. One speaker assumes the role of a "task manager" who must interact with the other speaker (the Wizard of Oz) who alone knows the conditions and rules of the task to be completed. Dialogues ranged from 40 seconds to 13 minutes in length.

Although not phonetically transcribed, TRAINS 91 includes orthographic transcriptions of the variations (e.g., *ohhh*) of base forms (*e.g., oh*) which approximate the token articulation of the given interjection. The transcription also includes overlapping speech. We semi-automatically tagged the non-lexical interjections and annotated them with the grounding moves of the DiSPEL taxonomy. We also noted the interjections' locations: that is, where in a turn they occurred (a *turn* being a speaker's contribution, which may include more than one utterance).

The HCRC Map Task corpus consists of collaborative problem-solving task dialogues between pairs of speakers. Each speaker had maps with slightly different landmarks (neither participant was able to see the other's map). One of the pair would instruct the other towards a particular destination (Boyle 1990). The HCRC Map Task corpus consists of 128 digitally-recorded dialogues spoken by a gender-balanced pool of 64 Glaswegian English speakers.

Although the orthographic transcription of Map Task did not reflect the token-type articulation of each non-lexical interjection, its transcriptions were time-aligned to audio files. With these, we did a broad phonetic transcription of the non-lexical interjections which occurred. The corpus also provided various codings including dialogue move and part-of-speech annotations that facilitated automated processing of data. The dialogue moves used in the Map Task annotations were not entirely grounding moves; thus, it would have been inappropriate to use them directly in our study. We compensated by automatically mapping them to the DiSPEL moves by also taking into account their part-of-speech annotations.

For our tests, we classed interjections first according to the discourse move taxonomy and then distinguished between those that function statically or dynamically in the discourse. We tested our hypotheses in two dialects of English.

6. The corpora: Choosing and post-coding

Finding appropriately annotated public-domain corpora with consistently annotated non-lexical interjections was difficult because early research in spoken language systems filtered out interjections as irrelevant to the process of communication. This problem restricted our corpus choice to the TRAINS 91 dialogues (Allen and Schubert 1991) and the HCRC Map Task Corpus (Boyle 1990, Carletta et al. 1996).

The TRAINS 91 corpus is a collection of 16 task-oriented Wizard of Oz game dialogues among gender-balanced and regional-dialect-balanced American English speakers. One speaker assumes the role of a "task manager" who must interact with the other speaker (the Wizard of Oz) who alone knows the conditions and rules of the task to be completed. Dialogues ranged from 40 seconds to 13 minutes in length.

Although not phonetically transcribed, TRAINS 91 includes orthographic transcriptions of the variations (e.g., *ohhh*) of base forms (*e.g., oh*) which approximate the token articulation of the given interjection. The transcription also includes overlapping speech. We semi-automatically tagged the non-lexical interjections and annotated them with the grounding moves of the DiSPEL taxonomy. We also noted the interjections' locations: that is, where in a turn they occurred (a *turn* being a speaker's contribution, which may include more than one utterance).

The HCRC Map Task corpus consists of collaborative problem-solving task dialogues between pairs of speakers. Each speaker had maps with slightly different landmarks (neither participant was able to see the other's map). One of the pair would instruct the other towards a particular destination (Boyle 1990). The HCRC Map Task corpus consists of 128 digitally-recorded dialogues spoken by a gender-balanced pool of 64 Glaswegian English speakers.

Although the orthographic transcription of Map Task did not reflect the token-type articulation of each non-lexical interjection, its transcriptions were time-aligned to audio files. With these, we did a broad phonetic transcription of the non-lexical interjections which occurred. The corpus also provided various codings including dialogue move and part-of-speech annotations that facilitated automated processing of data. The dialogue moves used in the Map Task annotations were not entirely grounding moves; thus, it would have been inappropriate to use them directly in our study. We compensated by automatically mapping them to the DiSPEL moves by also taking into account their part-of-speech annotations.

7. Markedness and unmarkedness

Our hypotheses lead to another sub-problem: we need to identify and define *unmarked* and *marked* sounds. Like discourse particles, the concept and definitions of markedness have a long reputation of controversy. Here, we are sympathetic to Greenberg's reinterpretation of this Prague School of Linguistics concept (Greenberg 1966, 1975).

In the Prague School of Linguistics, *markedness* referred to features that contribute to phonemic oppositions; it was what distinguished one phoneme from another. Hence markedness can be language-specific. Greenberg explicitly argued for a more functional or usage-based approach: markedness can have motivated cross-linguistic similarities. Items that are the most frequent, the simplest, the most universal, acquired the earliest by children, developed earliest in the language and contrast least with other unmarked categories are often considered *unmarked*. For us, this characterization implies that markedness can have cross-linguistic similarities as well as be language and even dialect-specific. Some perceptual and production advantages are fundamental to physiological capabilities (linguistically universal), whereas others may be conditioned to the contrasts present (the needs of) specific languages and dialects. Speakers of one dialect, for example, can maintain categorization (perceptual) boundaries and phonetic variations in articulation that differ from those maintained by speakers of another dialect. As such differences can exist from dialect to dialect, we had to identify the marked sounds relative to both American English and Glaswegian English.

Specifically for our task, we define *markedness* as characterized by the increased effort required for production and/or by its relatively greater perceptual salience. We looked at markedness only at the phone and syllable levels. We defined as *unmarked* any sounds that:

- *Occur in neutral or rest positions before or between articulations.* This is predominantly a physiologically-based criterion and should show cross-linguistic similarity.
- *Are typical sounds occurring in unstressed syllables.* Unstressed syllables, as explained with principles of PHB and as implied in psycholinguistic experiments (for example, Jusczyk et al. 1993), carry a low communication load. Since they convey relatively less information than stressed syllables, listeners do not need to attend as much to them. Less attention paid to unstressed syllables also means that less effort is needed. On this basis, unstressed syllables should involve sounds which are not as distinctive perceptually or production-wise. As phonological systems change from language to language, this criterion uncovers primarily language-specific and even dialect-specific differences.

– *Have low syllable complexity.* Syllable complexity can involve both cross-linguistic similarities and language-/dialect-specific differences. However, we only looked at cross-linguistic similarities in measuring syllable complexity by the presence of consonant clusters and the number of syllables. For example monosyllables, compared to multi-syllabic items, require less effort and are less salient perceptually. They are thus unmarked.

These criteria are not meant to be exhaustive. For example, non-average durations and pitch are also likely to indicate marked sounds, but analyses of these prosodic qualities and markedness on super-syllable levels remain to be investigated.

Given the criteria above, we reviewed the sounds that would be marked/unmarked for American and Glaswegian English. The lips when stably closed seem physiologically the most neutral non-articulating static position between utterances. That is, the resting articulatory position seems to correspond to /m/ without phonation and appears to be linguistically universal (Klaus Kohler 2003, personal communication). We therefore classed /m/ as unmarked for both American and Glaswegian English. As markedness is not always linguistically universal, we also classified markedness of sounds relative to each dialect.

The schwa is the most central position for an American English speaker, as well as a very typical vowel in unstressed and reduced syllables. Thus, for our analyses, we classed /m/ and /ə/ as the unmarked sounds of American English. Marked sounds then include rounded vowels (e.g. /o/), lengthened vocalizations (long vowels or mmmmm), non-central or tense vowels, and non-sonorants (such as stops).

For Glaswegian English, we looked to Abercrombie (1979) for the typical vowels it had in unstressed syllables. Figure 2 presents the typical vowels for Standard Scottish English (SSE) and the counterpart pronunciation in standard British English (Received Pronunciation, RP).

RP	SSE	Example
ə	ʌ	*china*
ɪ	ɪ	*father*
		pitted
	e	*pitied*

Figure 2. SSE Vowels typical in unstressed syllables (Abercombie 1979)

For both dialects, multi-syllables and consonant clusters were considered marked.

8. Results

The ten dialogues of TRAINS 91 included 410 tokens of non-lexical interjections. We analyzed six of the eight dialogues in the HCRC Map Task corpus, thus processing 3175 tokens of non-lexical interjections. For the TRAINS corpus, we first identified base forms by their relative high frequencies in contrastive functional realizations (as based on the taxonomy in Figure 1) and their locations in turns. In correlating the variations of the interjections to their base forms, articulatory or sound similarities are insufficient criteria because interjections with different functional realizations often have close sound structures.

We found that the patterns of frequencies in functional realizations, in addition to sound proximity, provided a reliable method of identifying variants of base forms even though frequencies were very low for some forms. As can be seen in Figure 3, in some cases we have left very low-frequency items (in parentheses) as base forms because of the possibility that they have functions that may be distinctive enough to warrant their status as base forms. The intonation and other prosodic features will most likely be key to aiding this distinction. These issues will require further investigation and are out of the scope of this chapter.

Base-Forms	Variants	Moves	Locations
ah	(hmm)	Ack, Exp, Corr	2,1
(aha)		Corr	2
(eh)		Corr / Exp	1 / 2
(err)		Corr, Exp	2, 1
m-hm		Corr	0, 1
uh	*uhm*, uhh, (uhhm), (uhhh), (uhmm)	Exp, Corr, Ack	2, 1, 3, 0
um	*uhm*, umm, (uumm), (*uhhm*)	Exp, Ack, Corr	2, 1, 3, 0
uh-huh		Corr	0, 2
hm	(*hmm*), (m), (mm)	Exp, Ack	2, 1
oh	(o), (ohh), (oo), (ooh), (oooh)	Corr, Ack, Exp	1, 2, 3
(oops)	(whoops)	Corr	0, 1 / 2
(ouch)	(uch)	Corr	0
wow		Corr	1

Figure 3. TRAINS corpus: non-lexical interjections and their variants (see text for explanation)

In Figure 3, interjections and their variants are listed in decreasing order of frequency. Items in parentheses indicate very low total frequency (across base forms if a variant). Items in italics are sound synonyms, that is, variants that occur with different baseforms. Location indicates turn information: 0=constitutes turn, 1=at the beginning of turn, 2=within turn, 3=at the end of the turn.

Our data provide some support that interjections are context-dependent and that their function depends on a combination of their position in turns, their denotation, and the context in which they appear.

Looking at the functional distribution of the non-lexical interjections, the first observation is that their most frequent position is within the body of the turn ("2"). However, most of these were functionally self-expansion interjections (Exp), which indicate the current speaker's intention to further expand the utterance by contributing more information. The least frequent position is at the end of the turn ("3"). Therefore, in general, interjections appear to prepare the listener in predicting the following utterances. The 2% that occur at the end are predominantly interjections which speakers use for self-expansion (indicating an intended beginning of a turn) but were interrupted by the listener.

The second most frequent function of interjections is to indicate change or correction. The change could be topical, indicating an initiation of a new topic. Or it could be an indication of a speaker's knowledge update. These are all tagged as self-repair or self-realization (Corr, self). Continuing the general trend of location distributions, this type of interjection tended to appear in the body of the turn ("2"). The least frequent function is that of acknowledgment (Ack).

In support of our hypotheses, results show that the degree of markedness in the sound makeup of interjections relates directly to the degree of interaction (see Figure 4); this is as we hypothesized. The data also confirm that the syllable structure of static-functioning interjections tends away from multi-syllabic forms (0.5% multi-syllables) more than the dynamic-functioning ones (3.4% multi-syllables).

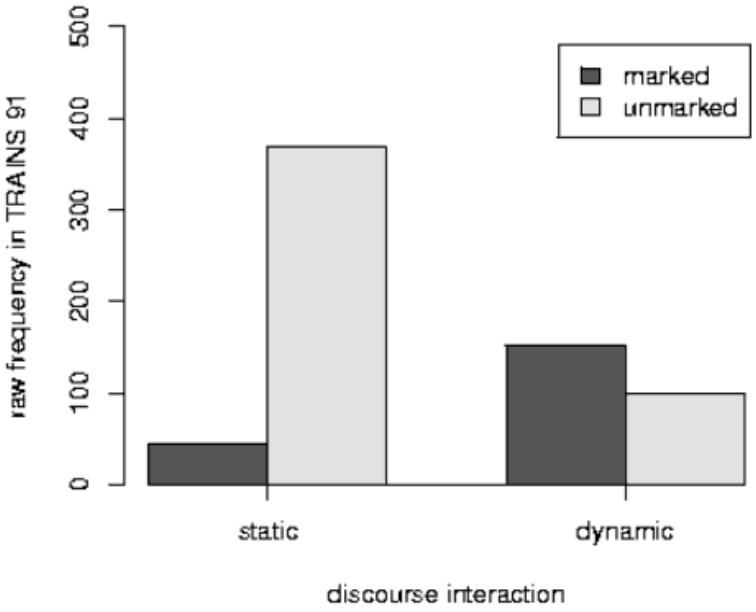

Figure 4. TRAINS 91: non-lexical interjection structure vs. discourse dynamics

The only sound which appeared not to match our predictions is the lengthened /m/; we had assumed that it is marked, yet it functions statically more frequently. However, a difference does exist between lengthened /m/ and its shorter form. The lengthened /m/ occurs primarily at the beginning and the end of turns (thus marking the change in turns) whereas the shorter form occurs primarily within turns. This may imply that the sound structure of non-lexical interjections depends on both function and location and supports our hypothesis that marked sounds indicate the dynamics of interactions.

9. Interjections in the HCRC Map Task corpus

Even a first pass of the Map Task analysis (in Figure 5) attests to our supposition that sound markedness is significantly interrelated with discourse dynamics.

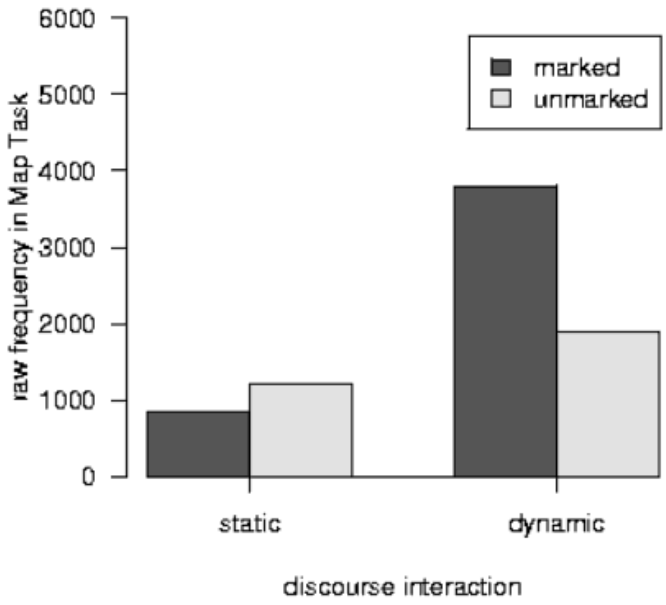

Figure 5. Map Task: non-lexical interjection sound structure vs. discourse dynamics

The results for the Map Task corpus are slightly more complicated than for TRAINS as there were many more multi-syllabic forms in the Map Task corpus (probably due to the nature of the task). Whereas the TRAINS analyses involve markedness that was primarily at the phone level, the Map Task results also reflect markedness on the super-phone level.

To examine the effect of syllable complexity as a criterion of markedness, we split the data in Figure 6 to show that taking into account the number of syllables also provides support for our hypothesis that static-functioning interjections too have unmarked sound structures and further strengthens our claims for sound-function interdependency.

Splitting the data according to syllabicity reveals that all multi-syllabic non-lexical interjections in Map Task have dynamic functions. This may indicate a stronger support for our hypotheses; however, this may also only mean that the HCRC Map Task corpus did not have multi-syllabic non-lexical interjections which functioned statically because of the nature of the task used during data collection. Nonetheless, the role of number of syllables becomes more convincing with a detailed look at a subset of the data.

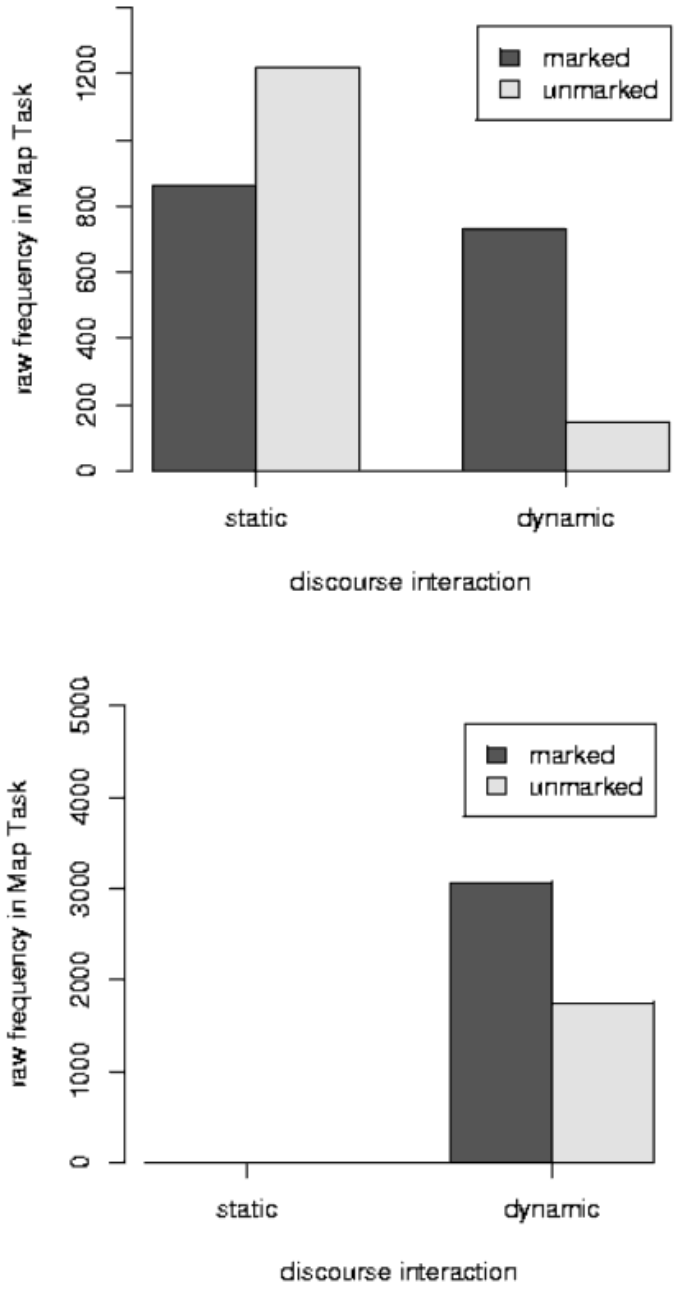

Figure 6. Map Task: syllable complexity in monosyllabic interjections (top) and in multisyllabic interjections (bottom)

Figure 7 presents the data for only the unmarked sounds in Standard Scottish English. Contrary to our hypotheses, the unmarked sounds /m/ and /ʌ/ seem to occur in dynamic-functioning items instead.

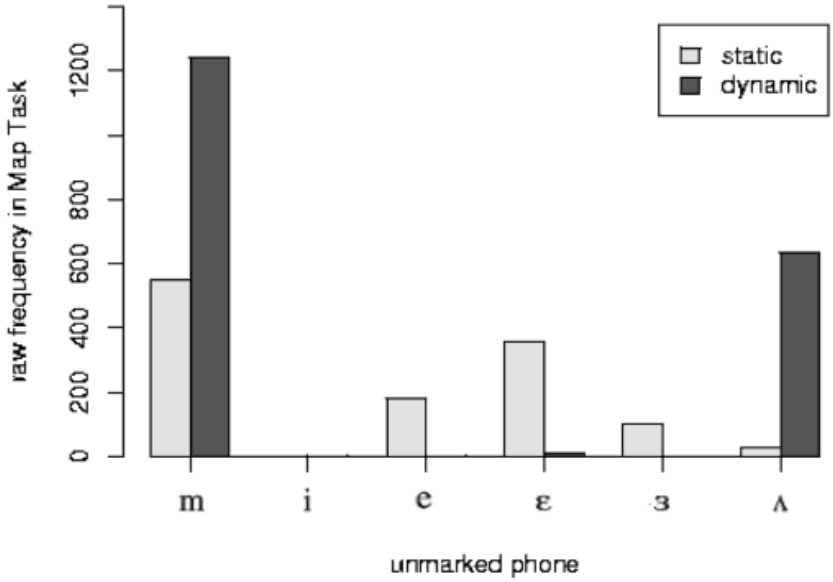

Figure 7. Map Task: markedness by discourse dynamics for unmarked phones

However, when we take into account syllable complexity, the data fit our hypotheses again. All the instances in which the unmarked /m/ had the "unexpected" occurrence in dynamic functions were because they occurred in bisyllabic items. Over half of the deviations involving the unmarked /ʌ/ can also be explained in the same way (see Figure 8): it is mostly dynamic when multisyllabic and less so when monosyllabic.

Analyzing the remaining deviant /ʌ/-dynamic function occurrences, we found that almost all of these instances were characterized by extremely short durations (average duration of 85 msec with a standard deviation of 26) and orthographically transcribed as *oh*. This might be that the target was /o/ but the sequence was too short for the acoustics to be perceived as the intended articulation, and it also indicates that duration analysis, as well as other prosodic investigations, may be insightful.

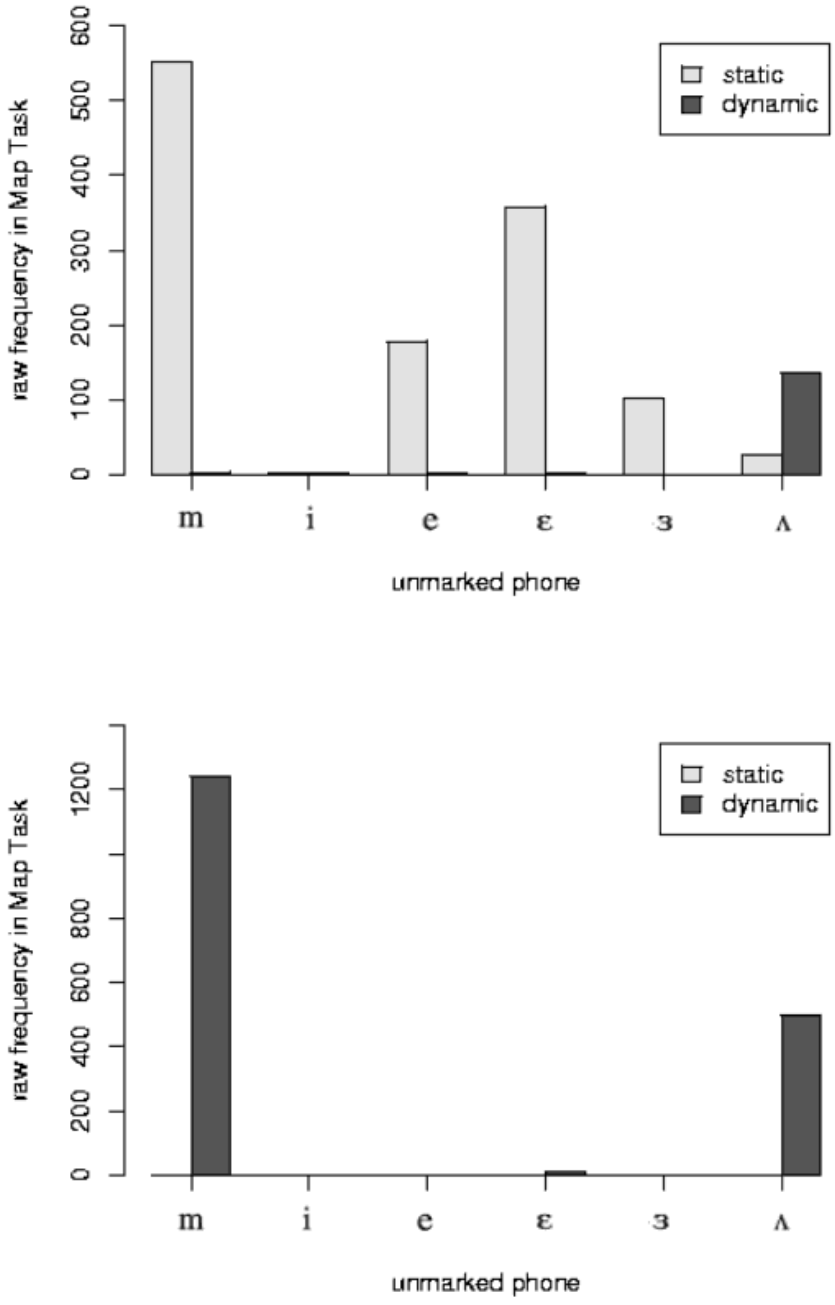

Figure 8. Map Task: effects of syllable complexity for unmarked phones: monosyllabic (top) and multi-syllabic (bottom)

10. Conclusions

The goal of our analyses was multi-fold. Our main pursuit was to find support for our claims of the importance of non-lexical interjections in linguistic analyses and to uncover their regularities. Following the corpus linguistic method of Phonology as Human Behavior, we tagged our non-lexical interjection data into factors that reflect the importance of the role of communication (in our case, as reflected in the discourse structure) and human constraints in communication (sound markedness). With these classifications we are able to find sound structure in non-lexical interjections which does not differ from the phonological principles of the dialect in which they occur. That is, like other sounds in a language, non-lexical interjections reflect the balance of human constraints in communication. Perhaps non-lexical interjections reflect this balancing act more transparently than other more lexicalized items.

Similar studies across different languages and their dialects would certainly be necessary, but the results are promising and reinforce the cognitive linguistic tenets that

- the apparent "noise" in language can actually be one of the most interesting and insightful windows into language phenomena;
- meaning, form and function are inseparable; and
- structure emerges when considering the factors and context impinging on language use.

Acknowledgements

We would like to thank our editors and reviewers for their invaluable comments. The ideas which Yishai Tobin and Liz Shriberg shared with us were important, but above all, this work could not have developed further without their encouragement. Thanks also to Matthew Aylett for help with the HCRC Map Task corpus.

References

Abelin, Asa. 1999. *Studies in Sound Symbolism*. Sweden: Goteborg University Dissertation. Gotehnburg Monographs in Linguistics.

Abercrombie, David. 1979. "Languages of Scotland." In *The Accents of Standard English in Scotland*. Edinburgh: W & Chambers, 68–84.

Allen, James F., and Lenhart K. Schubert. 1991. *The TRAINS Project*. [TRAINS Technical Note 91–1]. Department of Computer Science, University of Rochester, New York.

Boyle, Elizabeth. 1990. *User's Guide to the HCRC Dialogue Database*. [Technical Report HCRC/TR–11]. Human Communication Research Centre, University of Edinburgh.

Carletta, Jean, Amy Isard, Stephen Isard, Jacqueline Kowtko, Gwyneth Doherty-Sneddon, and Anne Anderson. 1996. *HCRC Dialogue Structure Coding Manual*. [Technical Report HCRC/TR–82] HCRC, University of Edinburgh.

Clark, Herbert H., and Jean E. Fox Tree. 2002. "Using uh and um in Spontaneous Speaking." *Cognition* 84: 73–111.

Clark, Herbert H., and Edward F. Schaefer. 1989. "Contributing to Discourse." *Cognitive Science* 13: 259–294.

Davis, Joseph. 1984 [1987]. "A Combinatory Phonology of Italian, Part I: Initial constriction to maximal aperture." In *Columbia University Working Papers in Linguistics* 8: 1–99.

Diver, William. 1979. "Phonology as Human Behavior." In Doris Aaronson and Robert W. Rieber (eds.), *Psycholinguistics Research: Implications and applications*. Hillsdale, N.J.: Lawrence Erlbaum, 161–186.

Diver, William. 1995. "Theory." In Ellen Contini-Morava and Barbara Sussman Goldberg (eds.), *Meaning as Explanation: Advances in linguistic sign theory*. Berlin: Mouton de Gruyer, 43–114.

Fischer, Kerstin. 2000. *From Cognitive Semantics to Lexical Pragmatics: The functional polysemy of discourse particles*. Berlin: Mouton de Gruyter.

Fox Tree, Jean E., and Josef C. Schrock. 1999. "Discourse Markers in Spontaneous Speech: Oh what a difference an oh makes." *Journal of Memory and Language* 40 (2): 280–295.

Fraser, Bruce. 1999. "What Are Discourse Markers?" *Journal of Pragmatics* 31 (7): 931–952.

Goffman, Erving. 1981. *Forms of Talk*. Oxford: Blackwell.

Greenberg, Joseph Harold. 1966. *Language Universals with Special Reference to Feature Hierarchies*. Berlin: Mouton.

Greenberg, Joseph Harold. 1975. "Research on Language Universals." *Annual Review of Anthropology* 4: 75–94.

Heeman, Peter A., Donna Byron, and James F. Allen. 1998. "Identifying Discourse Markers in Spoken Dialog." In *AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*, Stanford, California, 44–51.

Heeman, Peter A., and Donna K. Byron. 1998. *Discourse Marker Use in Task-oriented Spoken Dialog*. [Technical Report 664]. Department of Computer Science, University of Rochester.

Hirschberg, Julia, and Diane Litman. 1993. "Empirical Studies on the Disambiguation of Cue Phrases." *Computational Linguistics* 19 (3): 501–530.

Jusczyk, Peter W., Anne Cutler, and Nancy J. Redanz. 1993. "Infants' Preference for the Predominant Stress Patterns of English Words." *Child Development* 64: 675–687.

Kim, Sung-A. 1995. "Coronal/Non-Coronal Asymmetry in Place Assimilation." *Papers in Phonetics & Phonology* 35: 153–172. Austin, Texas: Texas Linguistic Forum, University of Texas.

Lickley, Robin. 2001. "Dialogue Moves and Disfluency Rates." *Proceedings of DISS* 01: 93–96, Edinburgh.

Montes, Rosa Graciela. 1999. "The Development of Discourse Markers in Spanish Interjections." *Journal of Pragmatics* 31: 1289–1319.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1972. *A Grammar of Contemporary English*. New York: Seminar Press.

Schegloff, Emanuel A. 1982. "Discourse as an Interactional Achievement: Some uses of 'uh huh' and other things that come between sentences." In Deborah Tannen (ed.), *Analyzing Discourse: Text and talk*. Washington, D.C.: Georgetown University Press, 71–93.

Schiffrin, Deborah. 1987. *Discourse Markers*. Cambridge, UK: Cambridge University Press.

Schiffrin, Deborah. 2001. "Discourse Markers: Language, meaning, and context." In Deborah Schiffrin, Deborah Tannen, and Heidi Ehrenberger Hamilton (eds.), *The Handbook of Discourse Analysis*, Malden, Mass.: Blackwell, 54–75.

Shriberg, Elizabeth E. 1994. *Preliminaries to a Theory of Speech Disfluencies*. PhD thesis, University of California at Berkeley.

Tobin, Yishai. 1997. *Phonology as Human Behavior: Theoretical implications and clinical applications*. Durham, N.C./London: Duke University Press.

Traum, David R. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. thesis, University of Rochester, New York.

Ward, Nigel. 1998. "The Relationship between Sound and Meaning in Japanese Back-channel Grunts." In *The 4*th *Meeting of the (Japanese) Association for Natural Language Processing*, 464–467.

Ward, Nigel. 2000. "The Challenge of Non-Lexical Speech Sounds." In *International Conference on Spoken Language Processing*, II: 571–574.

Ward, Nigel. 2006. "Non-Lexical Conversational Sounds in American English." *Pragmatics and Cognition*, 113–184.