Rethinking marginality: Interjections as the beating heart of language

Oxford linguist Max Müller once pontificated that “Language begins where interjections end”. Work in pragmatics turns this view on its head by studying language in its natural habitat of face-to-face interaction, where interjections help us every moment to calibrate understanding and use complex language efficiently. A guiding hypothesis for this panel is that at least some interjections are highly adaptive communicative tools, culturally evolved for the job of keeping our social interactional machinery in good repair (Yngve, 1970; Dingemanse 2017). Far from being marginal grunts, words like ‘oh!’, ‘mm’, ‘um’ and ‘huh?’ play central roles in the most sophisticated uses of language. As metacommunicative signals, they are one of the places where theories of mind and pragmatic reasoning come to the surface, and they afford human language a degree of flexibility, robustness and error-tolerance unmatched in other known communication systems.

This session brings together new research on the centrality of pragmatic interjections in language, with a special focus on items and interactional practices that play crucial roles in managing the back and forth of everyday interaction. These phenomena have been studied in disparate disciplines, as seen by the proliferation of available labels, including back channels, discourse markers, phatic interjections, collateral signals, response tokens and non-lexical conversational sounds. In this lies both a challenge and an opportunity. The challenge is to formulate a unified perspective that can provide conceptual foundations and ensure cumulative progress. The opportunity lies in the disciplinary diversity, which provides us with complementary methods that can deliver converging evidence on open questions.

Topics covered in the session include: the central roles of ‘marginal’ items in the pragmatics of human interaction; their linguistic status as lexical or nonlexical items; their multimodal composition, as items combining verbal and visual cues; their semiotic status, combining indexical, iconic and symbolic properties; their cross-linguistic attestations, including patterns of universality and diversity; the paths of semantic and pragmatic change leading to and from them; and their implementation in models of language processing, dialogue systems and conversational agents.

Tuesday June 11, room TU107, 13:30-17:00

1330-1400 Intro | Negotiating mutual understanding: a comparative and experimental approach
Marlou Rasenberg & Mark Dingemanse

1400-1430 Interjection as coordination device: feedback relevance spaces
Christine Howes & Arash Eshghi

1430-1500 Probabilistic Pragmatic Inference of Communicative Feedback Meaning
Hendrik Buschmeier & Stefan Kopp

1500-1530 —break (30min)—

1530-1600 Turn structure & interjections
Christoph Rühlemann

1600-1630 Hebrew clicks: From the periphery of language to the heart of grammar
Yotam Ben Moshe & Yael Maschler

1630-1700 Interjections in Action
Isabel Ward & Nigel Ward

Evening: dinner w/ workshop participants
1 Negotiating mutual understanding in multimodal interaction: a comparative and experimental approach

Mark Dingemanse¹² & Marlou Rasenberg¹²
¹Radboud University ²Max Planck Institute for Psycholinguistics

In social interaction, information is never simply imparted: people work hard to co-construct meaning, negotiate mutual understanding, and reach conceptual alignment. To enable this metacommunicative work, natural languages feature a powerful set of interactional resources besides the content words and grammatical structures that have long been the staple of general linguistics. Here we report on work that investigates these resources from two methodological perspectives: (i) pragmatic typology, in which we compare conversational corpus data across languages to understand what shapes metacommunicative resources; (ii) experimental tasks, in which pairs of people solve several rounds of coordination problems to study how mutual understanding is built bit by bit in interaction.

The comparative study builds on prior work that has shown the universal relevance of interactional resources like backchannels, change-of-state tokens and other-initiated repair. For repair, even the shape of the interjection ('huh? [ã?]) has been found to be highly similar across languages, thought to be a result of convergent cultural evolution. The primary object of the work reported in this study is to test the hypothesis that this kind of convergence is also found for other interactional interjections (e.g. 'm-hm', 'oh'), and extends to nonverbal features. A comparison of a sample of 10 unrelated languages is expected to reveal a picture of constrained diversity. Pilot data indicates that every language likely has at least some forms that conform to what appear to be a common cross-linguistic template, but there is quite some room for a language-specific flowering of resources (as in Japanese backchannels).

The experimental study looks at verbal and visual resources for grounding and repair in a director-matcher task with switching roles. The task requires people to communicate about 16 unfamiliar objects in 6 consecutive rounds, enabling us to study the role of processes of grounding and repair in the formation of multimodal conceptual pacts. We expect to find an interplay of relatively low-level processes of automatic alignment and more high-level metacommunicative resources. Pilot data reveals that co-speech gestures can operate to silently make representations more convergent, while repair is used to explicitly revise conceptual pacts and backchannels to affirm them.

Putting the two perspectives together, we expect that crosslinguistic variation uncovered in study 1 is rich yet constrained, and that the experimental microcosm of interaction in study 2 may help us to get a grip on some of the factors that shape and constrain this variation.

2 Interjection as coordination device: feedback relevance spaces

Christine Howes\textsuperscript{1} & Arash Eshghi\textsuperscript{2}

\textsuperscript{1}University of Gothenburg \textsuperscript{2}Heriot-Watt University

Dialogue is co-constructed by multiple interlocutors with frequent feedback demonstrating whether something said is taken as understood [1,2]. To achieve this grounding, we produce relevant next turns or interjections (‘mm’, ‘yeah’). Some interjections indicate processing or coordination difficulties and the need for repair (‘huh?’). This feedback does not just occur at the ends of turns, but sub-sententially, showing that grounding occurs incrementally, before a complete proposition has been produced/processed [3,4].

However, despite evidence that speaker switch can occur anywhere, even within syntactic constituents [5], feedback is not appropriate just anywhere -- randomly placed backchannels disrupt the flow of dialogue, are rated as less natural and decrease rapport [6]. Using Dynamic Syntax [7], we provide a low-level, semantic processing model of where feedback ought to be licensed -- feedback relevance spaces (FRSs). These are analogous to (but more common than) transition relevance places (TRPs; [8]) -- places where speaker switch may occur. Just as this is optional at TRPs, feedback is optional at FRSs. The model accounts for cases where feedback occurs at FRSs, and also describes how it can be integrated at non-FRSs due to the predictive, incremental and interactive nature of Dynamic Syntax. In contrast to models of feedback that incorporate higher order reasoning about mental states [9], this model shows how feedback serves to continually realign processing contexts without recourse to higher order pragmatic reasoning, and provides a mechanistic model of the characteristic divergence and convergence that is key to moving dialogue forward.

As well as providing insights into human-human communication, this work has implications for the production and interpretation of human-like feedback in dialogue systems; not just based on unanalysed features (which may result in accurate placement), but because they have successfully compiled a semantic unit. Our FRS model is implemented [10] and deployed in a new dialogue system architecture [11]; ongoing work explores the naturalness and usability of such systems versus turn-based systems or those without the FRS model.

3 Probabilistic Pragmatic Inference of Communicative Feedback Meaning

Hendrik Buschmeier & Stefan Kopp
Bielefeld University

Communicative feedback is an expression of addressees’ listening-related mental states that parallels and influences their dialogue partners’ speech production (Clark 1996) by expressing ‘basic communication functions’ (e.g., perception, understanding, acceptance; Allwood et al. 1992). When occurring in the form of pragmatic interjections (e.g., ‘mm’, ‘huh?’) feedback occurs in a large number of forms. Applying phonologic, morphologic, or syntactic operations results in a combinatorially growing space of feedback expressions. These can be further varied using nonverbal markers (prosody, gesture; Freigang et al. 2017), which add continuous dimensions to the feedback form-space. Humans exploit this richness in form to enrich feedback meaning with attitudinal or epistemic components and to express subtle differences on various dimensions (e.g., certainty, degree of understanding, ongoing cognitive processing). Although the mapping between the form of feedback and its meaning has some aspects that are conventionalised, feedback meaning is idiosyncratic and relies heavily on iconic properties and – as a purely interactional phenomenon – on its dialogue context.

Because of this, we see communicative feedback as a ‘model phenomenon’ of language processing that allows for modelling the cognitive processes underlying pragmatic reasoning in language use without the need to model all of language. We present a computational model of feedback interpretation (Buschmeier 2018), which embodies a probabilistic approach to pragmatic inference (Goodman and Frank 2016) and conceptualises speakers’ feedback interpretation as attribution of listening-related mental states to their feedback-providing interlocutors. Given an addressee’s feedback and its dialogue context, the model attributes a second order belief-state to the addressee (a probability distribution over their listening-related mental states, such as perception, understanding, acceptance, etc.). The model is thus able to (1) represent and reason about a speaker’s degree of belief in the dimensions and grades of their listener’s listening-related mental states (e.g., there is a high probability that the listener’s understanding is estimated to be low). And (2) model the traditional semantic and pragmatic processes assumed to underly the hierarchical relationship of feedback functions (Allwood et al. 1992, Bunt 2011), namely ‘upward completion’ (Clark 1996) and ‘upper-bound implicata’ generated by the cooperative principle (Horn 2004).

We combined this model of feedback interpretation with an incrementally adaptive natural language generation model in an artificial conversational agent and evaluated it in a semi-autonomous Wizard-of-Oz study (Buschmeier 2018). Autonomously interpreting its human interlocutors’ multimodal feedback and adapting to their needs, this ‘attentive speaker agent’ communicated more efficiently than an agent that explicitly ensured participants’ understanding. Participants rated the agent more helpful and cooperative and found it to be able to understand their mental state of listening.

4 Turn structure and interjections

Christoph Rühlemann
Marburg University

The structure of turns is still poorly studied empirically. Sacks et al. (1974) noted that turns “regularly have a three-part structure: one which addresses the relation of a turn to a prior, one involved with what is occupying the turn, and one which addresses the relation of the turn to a succeeding one” (Sacks et al. 1974: 722). That is, a turn often does not start with its main business—the turn-constructional unit (TCU)—but rather with a pre-start, whose main function is to relate the current turn to the previous turn. Pre-starts have crucial sequential and interactional implications as they indicate the kind of movement the new speaker is going to make vis-à-vis the movement performed by the prior speaker be it “congruent with the understandings, expectations and projections of the previous turn” or as a departure “from some of these understandings, expectations and projections (Heritage 2015: 89). Pre-starts thus are early indicators whether, and how, the new turn will align with the trajectory for joint action set up in the prior turn and are also likely key signals in listeners’ processes of action ascription (cf. Levinson 2013; Levinson & Torreira 2015).

The aim in this talk is to examine Sacks et al.’s observation of the “three-part structure” (Sacks et al. 1974: 722) based on large amounts of representative empirical data, specifically 10-word turns (ten words being the average number of words in turns; cf. Rayson et al. 1997; Rühlemann 2018) extracted from the conversational subcorpus of the the British National Corpus (BNC) (cf. Hoffmann et al. 2008).

The analysis proceeds in three major steps. First, based on the BNC’s mark-up for interjections I investigate how interjections are positionally distributed in large amounts of turns of varied lengths. The analysis shows that interjections are overwhelmingly attracted to turn-first (and turn-second) positions, the likely location of pre-starts. Second, a subsample of 1,000 10-word turns is manually coded for items performing a pre-start function. This analysis discovers that 86% of all interjections in the whole subsample perform such a pre-start function: interjections have their habitat in pre-starts. The third step is to focus on one such pre-starting interjection, the pragmatic marker ‘well’ used to foreshadow disagreement with the bias of the preceding turn (Heritage 2013; 2015). In particular, I examine the phonetic design of ‘well’ in larger sequential contexts and present evidence to suggest that phonetic parameters such as duration, articulation, and pausing directly correlate with the level of disagreement expressed in the ‘well’-prefaced turn, with long, fully articulated, and pause-separated ‘well’ prefacing strong disagreement and short, reduced, and non-separated ‘well’ introducing weak disagreement. The marker ‘well’ thus showcases the remarkable versatility of interjections in implementing the ‘front-loading bias’ (Levinson 2013: 112), a fundamental bias toward inserting cues to action type early in the turn to facilitate action ascription.
5 Hebrew clicks: From the periphery of language to the heart of grammar

Yotam Ben Moshe & Yael Maschler
University of Haifa

Non-phonemic clicks are a prime example of linguistic marginalia: phenomena common throughout the languages of the world, that are ignored in linguistic research because they are assumed to be inconsequential (Dingemanse 2017: 195). That they are typologically unexceptional was demonstrated by Gil (2013), who found them across continents and language families; that they are ignored is evidenced by the dearth of references in Gil’s work, which is based mostly on personal communication; they are marginal, he explains, since they lie outside ordinary phonemic inventories, have limited semantics, and are grammatically unintegrated.

Recent research reveals that the semantics of clicks in actual conversation run a range from the truly marginal, being mechanical byproducts of ‘gearing up to speak’ (Ogden 2013), to more linguistic and interactional functions such as stance-taking (ibid.), ‘new sequence indexing’ (Wright 2011), word searches (Wright 2005), and negation (González Temer 2014).

Using the empirical methodology of Interactional Linguistics (Selting and Couper-Kuhlen 2001, Couper-Kuhlen and Selting 2017), we analyze the clicks in naturalistic audio and video recordings from the Haifa Corpus of Spoken Hebrew (Maschler et al. 2017). The uses of clicks in Hebrew are shown to be even more varied, and more closely integrated with the grammar, including two previously undescribed functions in addition to those mentioned above. First, Hebrew clicks may function as repair initiators in self/other-initiated self/other-repair (Schegloff, Jefferson & Sacks 1977), as in the following token of self-initiated self-repair:

```
66 Ora:    ..hi--. sganit menahelet bank!
           "she’s... a deputy manager at a bank!"
67            ....tsk
68            ...snif.
           "a branch [of one]."
```

Another previously undescribed function of clicks is prefacing disaligned turns, as in:

```
212 Naama:    gam ?oti šixnat,
              "you’ve convinced me too,"
213           rlehoci ?otax.Ɂ
              "to send you [on the trip]."
214 Natalie:  Ltsk
215          Lʔani ʔomeret ʔet ze,
              "I say it,”
216          kol kax bebitaxon ʔacmi,
              "with such self-confidence,”
217 Maggie:   fʰ
218 Natalie:  L..šeze
              "that it”
219          ..ze-- šeker,
              "it’s a lie,”
220          ..ken?,
              "yeah?,"
```

Natalie has been telling her friends about an interview she passed thanks to being so “good at bullshitting.” Here Naama compliments her for being genuinely convincing (212-213). Natalie responds that the secret is her confident delivery (215-216); but the prefacing click (214) marks her upcoming turn as disaligned, indicating that this isn’t intended as bragging but as self-deprecating rejection of the compliment. Indeed, she subsequently clarifies that her confidence is fake (218-220), making her undeserving of success.
We argue that this, as well as several other uses of Hebrew clicks, qualify as prototypical discourse markers, because they function metalingually and meet also the structural requirement in the definition of discourse markers (Maschler 2009: 17). This status attests such clicks’ high level of integration in the grammar. Their presence alongside nearly-automatic clicks indicating ‘gearing up to speak’ enables us to describe Hebrew clicks ranging from the truly para-linguistic to the truly grammatical, and to hypothesize the paths of grammaticization which could have led to their drift from the periphery of language to the heart of grammar.

6 Interjections in Action

Isabel Ward & Nigel Ward
University of Texas at El Paso

Interjections are vitally important in realtime coordination of action, but this has not previously been directly studied. We recorded audio, screen images, and keystrokes as they occurred as four pairs of American English speakers played a cooperative online game, Fireboy and Watergirl. In 40 minutes of interaction we found 117 interjections. We examined each in context and grouped them into categories, primarily based on similarities in the state of mind conveyed, considering also the state of gameplay, the phonetic and prosodic forms, and the intended effects on the other player. We identified nine categories.

First there were interjections that related to information state. These conveyed the degree of understanding, either of the game situation or of the other person’s explanation. These ranged from hm? (lack of understanding) through hm and uh-huh (partial understanding) to oh and ahh (complete understanding).

Next there were three types of affect bursts. One expressed a feeling of accomplishment after the speaker or the other player accomplished a tricky jump or completed a level, often with hnn. The second expressed disappointment, typically after the player’s character died, as with aww in creaky voice or other sounds with /a/. The third expressed negative feeling about a situation, marked with a plosive followed by a fricative, such as ksssh.

Then there were four that related to the nature of the current activity. Effort, occurring in moments where a player was in a difficult situation and concentrating on moving correctly, was marked by multiple short syllables, separated with glottal stops and synced with pressing the jump key, such as urh uh or hu un. Alarm, which occurred at moments of imminent disaster, was often expressed with loud repeated syllables with glottal fricatives such as uh uh uhh. Concern and tension sometimes occurred at such times, marked with sharp inhalations, usually with sss sounds. Feelings of relief and release of tension often took the form of breathy, slow exhalations, which sometimes segued into laughter.

Finally there were sound effects, such as fwaa and hoo. These often conveyed enjoyment, with a lack of tension and effort, for example when falling freely or being carried by the wind. These also seemed to convey something about the kind of motion, in an onomatopoeic way, for example, collisions with oof and running without obstacles with whee.

We note that these categories have some overlap, that some interjections are multifunctional, and that some shift their nature halfway through, reflecting changes in the game state. Orthogonally to these categories, most interjections serve to show continued interest and engagement in the game. Although few directly affect the behavior of the other player, they help the players be aware of each other’s mental state, current and intended actions, and possible outcomes.

This study has added to the inventory of known functions of interjections, and illustrated how specific phonetic and prosodic properties help convey specific meanings.