

The grammar of engagement I: framework and initial exemplification

NICHOLAS EVANS

Australian National University & ARC Centre of Excellence for the Dynamics of Language

HENRIK BERGQVIST

Stockholm University

AND

LILA SAN ROQUE

Radboud Universiteit Nijmegen & Max Planck Institute for Psycholinguistics

(Received 08 March 2017 – Revised 29 August 2017 – Accepted 31 August 2017)

ABSTRACT

Human language offers rich ways to track, compare, and engage the attentional and epistemic states of interlocutors. While this task is central to everyday communication, our knowledge of the cross-linguistic grammatical means that target such intersubjective coordination has remained basic. In two serialised papers, we introduce the term ‘engagement’ to refer to grammaticalised means for encoding the relative mental directedness of speaker and addressee towards an entity or state of affairs, and describe examples of engagement systems from around the world. Engagement systems express the speaker’s assumptions about the degree to which their attention or knowledge is shared (or not shared) by the addressee. Engagement categories can operate at the level of entities in the here-and-now (deixis), in the unfolding discourse (definiteness vs indefiniteness), entire event-depicting propositions (through markers with clausal scope), and even metapropositions (potentially scoping over evidential values). In this first paper, we introduce engagement and situate it with respect to existing work on intersubjectivity in language. We then explore the key role of deixis in coordinating attention and expressing engagement, moving through increasingly intercognitive deictic systems

Address for correspondence: Nicholas Evans. e-mail: nicholas.evans@anu.edu.au

from those that focus on the the location of the speaker, to those that encode the attentional state of the addressee.

KEYWORDS: engagement, attention, intersubjectivity, deixis, coordination.

These two individuals, the producer and the recipient of language, or as we may more conveniently call them, the speaker and the hearer, and their relations to one another, should never be lost sight of if we want to understand the nature of language and of that part of language which is dealt with in grammar.

(Jespersen, 1924, p. 17)

1. Introduction

As speakers of English or similar languages, we are prone to presume that the meaning-categories found in our grammar represent the essential information of a situation, and that whatever we express by more peripheral methods is just so much ‘extra stuff’. Take the English sentences *The sun is coming up, look!* or *Hey, the sun has come up!* (said as I shake my companion awake). In each of these, the core grammatical categories of tense, aspect, and mood dictate the choice of auxiliary and verbal inflection (*is coming up, has come up*). In contrast, expressions which position the speaker’s assessment of what information their interlocutor has access to lie at the sentence periphery and would not normally be seen as part of grammar. *Look!* presumes current non-access and directs attention. *Hey!* expresses dawning awareness – either speaker surprise, or directing the attention of a presumably non-aware addressee. Certainly, the choice of a phrase like *has come up* can indicate an assumption by the speaker that the described event is news to the interlocutor (McCawley, 1981; McCoard, 1978), but this meaning is only one of several that are available with the English perfect (Michaelis, 1994), rather than a dedicated, necessary meaning. And starting a sentence with a word like *certainly*, as we do above, may be one of many tools we use to both concede and coerce an addressee’s point of view, but it is hardly a core component of forming a grammatical English clause. But this division of labour in English and languages like it – with the grammar focusing on event structure, and pragmatic questions of intersubjective placement outsourced to more marginal parts of the system – has distorted our view of what grammar can do. In these two serialised papers, we ask the reader to accompany us in some typological gymnastics which will show that there are numerous languages which place such ‘pragmatic’ factors at the heart of their grammars, and give their speakers neat shortcuts for expressing complex and delicate matters of who knows (or could, or should, know) the situation or event that is being described.

Taking a step back from what is familiar in English and its congeners, it should come as no surprise that there are languages which place intersubjective

alignment at their heart. After all, grammars routinise our most common and central communicative tasks. And theory of mind (e.g., assessing an addressee's attentional state), and the ability to coordinate attention with others (e.g., through awareness of whether another has perceptual access to the same or different things as we do) are central and defining human skills, and underpin many elements of social cognition (e.g., Enfield & Levinson, 2006; Goody, 1995; Tomasello, 2008; Tomasello, Carpenter, Call, Behne, & Moll, 2005). Likewise, ostensive demonstration by adults, and children's subsequent directing of attention, are a key part of adult-child interactions and set the scene for 'natural pedagogy' that is unique to humans¹ and common to all cultures (Csibra & Gergely, 2009, 2011). The ability to achieve such primary intersubjectivity (Trevarthen, 1979; cf. Scott-Phillips, 2015; Sperber & Wilson, 1986) has been argued to be a prerequisite for the evolution of culture, and in particular of those conventionalised cultural manifestations which form linguistic signs.²

Achieving intersubjectivity thus lies at the heart of how human communication systems evolved. But beyond this, speakers in real time need constantly to bring about adjustments to each other's attention, beliefs, and states of knowledge – directing, persuading, and informing, at the same time as indicating empathy and deference (or their absence). Every human communicative system has a rich set of ways of doing this, many lying outside the domain of what is normally conceived of as grammatical structure. For example, Stivers and Rossano (2010) outline strategies used by speakers to mobilise the response of their addressee – gaze to the addressee, interrogative syntax, interrogative intonation, and speaking about topics that belong to the epistemic realm of the addressee. To this we might add gesture, and stance-taking phrases of various types (see Kockelman, 2004, Biber & Finegan, 1989). Detailed investigations of these communicative resources have been pursued in discourse analysis (e.g., Verhagen, 2005, 2015) and in the conversation-analytic tradition (Heritage, 2011, 2012a, 2012b; Sacks, 1987; Sacks, Schegloff, & Jefferson, 1974; Schegloff, 2007).

But, despite the centrality of this communicative task, our understanding of the full panoply of grammatical means used across languages for intersubjective coordination remains basic (see, e.g., Heritage's comments on the possible 'shortchanging' of linguistic form in his own work on epistemics in action; 2012c, p. 76). In this paper we return the focus to linguistic form, and in particular grammatical organisation. We argue that many languages have

[1] Though for debate on whether ostensive demonstration or attention-direction are indeed confined to humans, see Moore (2015).

[2] The coordination of attention and belief reasoning (cf. 'shared intentionality'; Tomasello, 2008, 2014) are further central to the debate concerning how theory of mind develops in the child and whether this development may be equated with one or several cognitive abilities (see Apperly & Butterfill, 2009, for a discussion).

grammaticalised systems for monitoring and adjusting intersubjective settings; it is this *grammaticalised intersubjectivity* which we refer to as *engagement*, in much the same way as grammaticalised time representation merits the special metalinguistic term *tense*.³ Our paper is serialised into two parts, across two successive issues of this journal – the first introducing the phenomenon, situating it with respect to other work on intersubjectivity in language, and outlining the key role of deixis in coordinating attention, the second broadening out to a typological survey of the phenomenon of engagement and to the diachronic question of how engagement systems originate.

Within this first part, we begin with an initial example from the Colombian language Andoke (§2), whose description by Landaburu (2007) was the first to argue for engagement as a core grammatical phenomenon. We then review two other bodies of work on epistemic distribution in the speech situation. The first research tradition (§3) is attuned to general properties of conversational organisation rather than the use of core grammatical devices. The second (§4) sets up a general framework for viewing multiple perspective in language, necessary to understand the asymmetries of knowledge distribution that accompany any projection by the speaker of what they believe (or wish to portray they believe) the addressee's epistemic disposition to be.⁴ In §5, the concluding section of Part I, we pass to the primal scenario for establishing shared access – deixis – and examine the notion of engagement as it applies to the management of joint attention in deictic scenarios of drawing attention to entities, through demonstrative systems such as those of Turkish and Jahai.

2. What is engagement? An initial example

Consider the following pair of contrasting sentences from Andoke, an isolate language of the Colombian Amazon (Landaburu, 2007).⁵

[3] While we may attribute our use of the term 'engagement' to Landaburu's work on Andoke, we also note that it has been used by others to discuss overlapping phenomena in discourse studies (e.g., Hyland, 2005) and in French linguistics, notably Desclés (2009) and Guentchéva (2011).

[4] As Alan Rumsey points out (p.c.), dissembling may be involved at various levels. How speakers use particular formal devices cannot be taken as a direct reflection of what they think or assume; often it is more a matter of their Goffmanian 'presentation of self' in particular situations. The speaker may be deceptive in the belief they ostensibly project about the addressee, as they may be about their own knowledge state. Caveats about these devices pertaining to 'presented belief' rather than actual belief thus need to be added. However, since adding these caveats at every relevant point in our discussion would clutter our exposition, we confine ourselves to stating it once here.

[5] Abbreviations: 1: first person, 2: second person, 3: third person, ADDR: addressee, AGT: agent, ASYM: asymmetric, DAT: dative, ENCL: enclosure, ENGAG: engagement, GER: gerundive, INAN: inanimate, INGR: ingressive, IQ: WH-question, NONMUTDEM: non-mutual demonstrative, PERV: perfective, PQ: polar question, SPKR: speaker.

- (1) a. *páa* *b-a* *λ-pó'kǎ-i*
 already +SPKR+ADDR.ENGAG-3SG.INAN 3SG.INAN-light-AGR
 'The day is dawning (as we can both see).'
- b. *páa* *kě-ø* *λ-pó'kǎ-i*
 already +SPKR-ADDR.ENGAG-3SG.INAN⁶ 3SG.INAN-light-AGR
 'The day is dawning (as I witness, but which you were not aware of).'

The relevant point of grammatical contrast is seen in the auxiliaries *b* and *kě* (structurally similar to a word such as *is* in the English phrase *is dawning*) that precede the main verb *λpó'kǎi* 'light(en), dawn'. The Andoke auxiliaries are made up of two parts: the first element (*b-* or *kě-*) encodes the dimension of 'engagement' – the relative access of speaker and hearer – and the second element marks subject agreement (i.e., who is undertaking the activity; in this case, the day or the sun itself, which is encoded as a third person singular inanimate subject). No descriptive sentence can be constructed without employing one element from the engagement set.⁷

Consider the situation where the day is dawning and the two of us, speaker and hearer, are watching the sun rise together, so the speaker can presume joint attention to this mutually accessible event. This would be expressed as in (1a), using the auxiliary base *b-* (represented as 'plus speaker and plus addressee engagement', +SPKR+ADDR.ENGAG). But if the event is not accessible to the addressee – for example, he is only just waking up and is not attending to it – the base, *kě-* ('+SPKR-ADDR.ENGAG') would be chosen (1b).⁸ Though the reference to 'seeing' in our elaborated translations may seem reminiscent of evidentials, in particular those marking the source of information as visual, what is at issue in examples like (1a, b) is not primarily the source of information but whether the addressee is presumed to be attending to, or more broadly to have access to, the event: pure evidentiality is about sources, whereas engagement is about the presumed

[6] The zero morpheme is given in the gloss of the original (Landaburu, 2007, p. 26) without explanation, but we presume it is a variant of the 3SG inanimate suffix.

[7] "Tipológicamente y conceptualmente es muy importante recalcar que no es posible formular una oración descriptiva sin escoger una de estas 4 marcas" (Landaburu, 2005). By *oración descriptiva* we take him to mean a non-imperative sentence, since both declarative and interrogative examples are found in these papers.

[8] For all Andoke examples the glosses are ours, in the spirit of Landaburu's own glossing but making explicit contrasts he sometimes only makes in accompanying tables. Most importantly, he gives the same gloss to all members of the engagement contrast set (EPISTEMICO in Landaburu, 2005, ENGAGEMENT in Landaburu, 2007), but assigns different values of speaker and hearer knowledge for the different forms in accompanying tables – we employ the values in his tables (as per our Table 1) in our glosses here. The minimal quintuplet assembled in (2) and (3) is compiled from two separate publications by Landaburu, neither of which gives the full set: Landaburu (2005) gives all but (3b), and Landaburu (2007) gives all but (3c).

presence or absence of intersubjective sharing, whatever the source. We will see later, however, that many languages exhibit complex interactions between engagement and evidentiality (Part II, §3).

As a second example, consider how one would translate ‘it’s the white people arriving’ into Andoke (Landaburu, 2007, p. 25). In a standard situation, with shared access to the event, the ‘shared engagement’ auxiliary base *b-* (2a) would be used – for example, where both the speaker and addressee are together in a canoe, the speaker hears the noise of a distant motor, and directs the addressee to pay attention to it, confident that they, too, will be able to hear it. On the other hand, the ‘unshared engagement’ auxiliary base in (2b) would be used in situations where (i) the interlocutor does not have direct access to the event described, but (ii) the speaker is sure of their assertion. A strong internal revelation to the speaker would be one such context; another would be the case where the speaker is up in a tree and from there sees the white people, whose arrival would not be visible to the addressee, positioned at the foot of a tree in the forest.

- (2) a. *duiʌhʌ* *b-ə̃* *dā-ə̃-ʌ*
whites +SPKR+ADDR.ENGAG-3PL INGR-move-3
‘It’s the whites arriving (as we can both witness).’
b. *duiʌhʌ* *kē-ə̃* *dā-ə̃-ʌ*
whites +SPKR-ADDR.ENGAG-3PL INGR-move-3
‘It’s the whites arriving (which I know / can witness but you can’t).’

This initial two-way contrast (shared accessibility versus speaker-only accessibility) is, in turn, part of a four-valued set of auxiliary bases (with a further subdivision of one value) whose other members deal with cases where the speaker *lacks* knowledge. In the case of true questions, where the interlocutor can be expected to know the answer, the pair *k-/d-* is used (Landaburu, 2007, p. 27): *k-* for polar (yes-no) questions such as ‘Is it the whites who are arriving?’ (3a), and *d-* for WH-questions like ‘Who is coming?’ (3b). The fourth value, coded by *bā-*, is used for self-interrogatory questions to which the speaker expects no answer from their interlocutor, who is simply a witness to the speaker’s deliberation; that is, the event is presented as inaccessible to both parties (3c).⁹

- (3) a. *duiʌhʌ* *k-ə̃* *dā-ə̃-ʌ*
whites -SPKR+ADDR.ENGAG.PQ-3PL INGR-move-3
‘Is that the whites arriving?’

[9] As a further example of the -SPKR-ADDR.ENGAG *bā-*, Landaburu (2007, p. 28) gives the example of an aged narrator, describing a genocide he witnessed as a child, using the form *bā-* as auxiliary base in the question ‘And why were they killing?’ Given the setting, in which the interlocutors were all too young to have witnessed the terrible events which he is recalling, Landaburu argues that this can only be self-interrogation, and that the addressees are not being expected to supply any type of answer.

- b. *kói* *d-ǎ* *dǎ-ǎ-α*
 who -SPKR+ADDR.ENGAG.IQ-3PL INGR-move-3
 ‘Who is arriving?’
- c. *duiɬa* *bǎ-ǎ* *dǎ-ǎ-α*
 whites -SPKR-ADDR.ENGAG.PQ-3PL INGR-move-3
 ‘I wonder if those are the whites coming’. (Landaburu, 2005, p. 2)

As Guentchéva and Landaburu (2007, p. 5) put it, the contrast between the auxiliary bases of Andoke “is better seen, not simply as a relation between the speaker and the truth of their statement but also ... as a relation between what the interlocutors know”.¹⁰ Further, Landaburu argues (2007, p. 30) that “as well as the knowledge of the speaker, we are dealing here with relations of epistemic authority between the speaker and the hearer. The speaker’s judgment of the truth of his proposition combines with the intersubjective dimension of the proposition, inside the grammatical system and not simply in perlocutionary or pragmatic effects.”¹¹

As Table 1 shows, Landaburu posits an orthogonal pairing of two two-valued semantic dimensions, neatly accounting for the functional symmetry of the Andoke system. (He treats *k-/d-* as specific variants conditioned by polar vs. WH-question as seen above.)

We adapt his terminology slightly in the translation process, substituting ‘knowledge’ vs. ‘lack of knowledge’ for his terms ‘savoir’ vs. ‘non-savoir’, and ‘speaker’ and ‘addressee’ for his ‘je’ vs. ‘tu’. In addition to these merely translational changes, we comment here on two more substantive problems of terminology. First, Landaburu’s terminology conceals a deep asymmetry: the speaker knows what they themselves know, but can only presume what the addressee knows, so that a more realistic characterisation of the terms in the left-hand column would be ‘presumed addressee (lack of) knowledge’, an issue we return to in §4 under the rubric ‘multiple perspective’. Second, neither Landaburu’s *savoir* nor its rough English equivalent ‘knowledge’ fully convey the range of the addressee’s mental dispositions: arguably, the crucial difference between the (a) and (b) example in each case concerns differential *accessibility* to the speaker and the addressee. In some of his examples it is clearly knowledge that is at issue, but in others, such as the ‘sunrise’ examples in (1), attention seems the more crucial mental disposition.

[10] “La fonction du préfixe gagne ainsi à être vue, non pas simplement comme un rapport de l’énonciateur à la vérité de son propos mais aussi ... comme un rapport entre les savoirs des interlocuteurs” (Guentchéva & Landaburu, 2007, p. 5).

[11] “Autant que du savoir du locuteur, il s’agit donc de rapports d’autorité épistémique entre le locuteur et l’interlocuteur. Le jugement du locuteur sur la véricité de son propos se combine avec la dimension intersubjective du propos, dans le système grammatical et pas simplement dans les effets perlocutoires ou pragmatiques” (Landaburu, 2007, p. 30).

TABLE 1. *The Andoke engagement paradigm as a 2 x 2 matrix (Landaburu, 2007, p. 30)*

	Speaker knowledge	Lack of speaker knowledge
Addressee knowledge	<i>b-</i>	<i>k-/d-</i>
Lack of addressee knowledge	<i>kẽ-</i>	<i>bã-</i>

Landaburu presciently observes (2007, pp. 30–31) that it was unlikely that the contrasts he described there would be found just in Andoke, and that further research would probably turn up comparable phenomena elsewhere. Moreover, he suggests that an emphasis on speaker-knowledge, at the expense of the epistemic relations between speaker and addressee, results from the influence of traditional grammar (whose assumptions were then imported into formal logic), itself reflecting the contingent privileging of certain grammatical categories (tense, aspect, mood) in the classical Indo-European languages.

There are, of course, important and familiar exceptions to the lack of attention paid to grammaticalised epistemic relations between speaker and hearer. The most important are (a) the definiteness contrasts expressed in article systems in western European languages,¹² (b) focus systems responsive to information structure,¹³ and (c) discourse particles¹⁴ like German *doch* ‘after all, actually

[12] “(W)hat type of shared knowledge is needed for language use? and ... how is that shared knowledge in practice assessed and secured? The area of language in which we will take up these questions is definite reference, but even our interest in definite reference is secondary to our concern with the two questions of mutual knowledge” (Clark & Marshall, 1981, p. 11).

[13] The linguistic term ‘focus’ is notoriously variable in its use, being generally partitioned into ‘referential givenness/newness’ and ‘relational givenness/newness’ (Gundel & Fretheim, 2006). The latter pertains to divisions of a linguistic unit into given/new, topic/focus, etc., and is not relevant to the phenomena discussed in this paper. The former is defined by Gundel and Fretheim as “a relation between a linguistic expression and a corresponding non-linguistic entity in the speaker/hearer’s mind, the discourse (model), or some real or possible world, depending on where the referents or corresponding meanings of these linguistic expressions are assumed to reside.” This is closer to many of the phenomena discussed in this paper, though we note the lack of precision with regard to whose mind is involved, or the nature of the intersubjective relationship between them. Elsewhere in the same paper they mention “the speaker/writer’s intention to affect the addressee’s attention state”. This draws their conception of focus closer to the typical purpose of engagement, as discussed in this paper, but the encoding devices they discuss are less grammaticalised and involve prosody and syntactic positioning.

[14] Significantly, Heritage (2012c, p. 77) states that “deep and important findings await us ... in an increasing body of cross-linguistic analyses of various epistemic particles (Hayano, 2011, 2012; Wu, 2004)” (see also Wu & Heritage, 2017). We briefly return to the particle issue in Part II, §5. For now, we simply note that while epistemic particles do indeed often encode the sorts of epistemic assessments we are interested in here, they differ from the prototypical systems of engagement in being less integrated into the grammar (e.g., as relating to their status as particles rather than affixes), and being less structured into symmetrical systems of opposition on more than one dimension.

(against earlier expectation)' or Italian *mica* 'not at all (against earlier positive expectation)' which express incompatibilities between an asserted state and that presumed to have been the case at some prior moment in the discourse.¹⁵ For many investigators of information structure, which takes in "such psychological phenomena as the speaker's hypotheses about the hearer's mental states" (Lambrecht, 1994, p. 3), it is a precondition that "what one individual may know or hypothesize about another individual's belief-state" is only of analytic interest "insofar as that knowledge and those hypotheses affect the forms and understanding of LINGUISTIC productions" (Prince, 1981, p. 233).

All of these studies, then, are relevant to the domain of intersubjective coordination. But as we will show, they represent only a fraction of the grammatical design space. With the wider typological sample we adduce, it is clear that the world's grammars attest a much wider set of intersubjectively relevant categories than has previously been suspected. The initial typological framework we propose here aims to set out a broad programme of typological research that systematises the great diversity of grammatical devices in the intersubjective domain, along the following two axes:

- (i) scope, be it semantic or syntactic (entity/location/referent, state of affairs/proposition, evidence/metaproposition),
- (ii) intersubjective distribution (epistemic authority can be speaker, addressee, neither, or both).

A note on terminology before we proceed. Rather than burden the overworked term *intersubjectivity* with one further use, we will follow Landaburu's lead in using the term *engagement* to refer to *a grammatical system for encoding the relative accessibility of an entity or state of affairs to the speaker and addressee*.¹⁶ This definition clearly relates to Du Bois' (2007, p. 144) notion of 'alignment', "the act of calibrating the relationship between two stances, and by implication between two stancetakers".¹⁷ But whereas his term is intended to be broadly

[15] Cf. Kirsner (2003) for the use of the Dutch particles *hoor* lit. 'hear' vs. *hè* 'isn't it?' with imperatives.

[16] One understanding of the word 'accessibility' is in reference to perceptual access, for example, something that is visible to a person is also directly 'accessible' to that person (cf. Tournadre & LaPolla, 2014). However, our use of the word is broader than this, in that we also understand it in terms of *mental* accessibility and in relation to 'having something in mind'. For example, under this latter reading, something that a person is attending to is highly accessible, because it is at the forefront of that person's mind. We can thus think of attention (and other mental dispositions) as a kind of (or even constraint on) accessibility, along with visibility, audibility, etc.

[17] In fact, a similar use of the term 'alignment' goes back beyond Du Bois to Erving Goffman, who used it at least as far back as his 1974 book *Frame analysis*. In his subsequent book *Forms of talk* (1981) he defines footing (rather sketchily) as "the alignment we take up to ourselves and the others present as expressed in the way we manage the production or reception of an utterance".

functional, we reserve engagement for grammaticalised systems, which are only one means of addressing the alignment problem. Likewise, while the term ‘stance’ has been employed in somewhat similar ways by various authors, it is generally used in a broadly functional way rather than focusing on grammaticalised systems: examples are Heritage’s (2012a, p. 6) definition of ‘epistemic stance’ as concerning “the moment-by-moment expression of [social] relationships, as managed through the design of turns at talk”, or Engelbretson’s (2007) more general definition of stance as expressing ‘a personal belief or attitude’ or ‘social value’.

Finally, a remark on the trajectory by which categories are ‘typologically detached’ from semantically related categories that they share expression with in many languages. In laying out their analyses, it is helpful for typologists to work with canonical, neatly cut-and-dried categories (Brown, Chumakina, & Corbett, 2013), so as to illustrate the dimensions of the design space with maximal clarity. But the relation of engagement to epistemic categories means that it borders on many more familiar linguistic categories: evidentiality, miratives, focus, mood, and modality.¹⁸

And much of the time actual languages run some of these dimensions together. This may arise through conventionalised polysemous extensions across categories, e.g., the well-known case of Turkish *-mİş*, used both for evidential categories and for miratives (Aksu-Koç & Slobin, 1986; Slobin & Aksu-Koç, 1982). Or it may come about by exploiting inferences from one type of interpretation to another, e.g., by applying hearsay evidentials to one’s own past behaviour to indicate ironical disbelief or lack of responsibility for one’s unconscious actions (see, e.g., Michael, 2012; Wilkins, 1986). Our general strategy, in unfolding the typological framework we develop

[18] The foundational if rather abstract definition of mood by Jakobson (1990 [1957]) as characterising PnEn/Ps “the relation between the narrated event and its participants with reference to the participants of the speech event”, may be charitably interpreted as subsuming engagement since we are talking about intersubjective relations between participants in the speech event with respect to the narrated event, though his actual examples did not touch on phenomena comparable to those we discuss here. Likewise, consider the following interesting and inclusive definition of modality by Timberlake (2007): “Modality is about alternatives – how we come to know and speak about the world, how the world came to be as it is, whether it might be other than it is, what needs to be done to the world to make it what we want. The alternatives are sorted out and evaluated by some sort of authority, often the speaker or, if not the speaker, some other participant or even another situation. Modality, then, is consideration of alternative realities mediated by an authority” (p. 315). This could only be stretched to cover engagement if we include attentional phenomena – ‘who knows about or attends to it’ – under the rubric of ‘how we come to know and speak about it’, and even then there is no overt focus on intersubjective calibration. Other definitions of modality fit even less well, e.g., the one by Nuyts (2006, p. 1) as “any kind of speaker modification of a state of affairs, even including dimensions such as tense and aspect ... qualifications of states of affairs” which deviates from our interests through its exclusive concentration on the speaker.

here, is to begin each major section with more clear-cut cases and then look at more complex and transitional ones.

3. Epistemic management in conversation

In a series of papers, John Heritage discusses the related notions of ‘epistemic status’, ‘epistemic stance’, ‘epistemic gradient’, and ‘territories of knowledge’ in an effort to account for the relation between sentence-type and communicative function, and how this is seen in the sequential unfolding of turns as a form of social action (Heritage, 2002, 2011, 2012a, 2012b, 2013; Heritage & Raymond, 2005, 2012). He argues that epistemic status and epistemic stance are keys to understanding the discrepancies between grammatical form and (social) action, an issue that has plagued speech-act theory since its formulation (Austin, 1962; Searle, 1969) and necessitated the label ‘indirect speech-acts’ to account for such discrepancies (see Levinson, 1979, 1983, for a critique).

Epistemic status, as an index of relative epistemic authority, is formulated with reference to the notion of A- and B-events (Labov & Fanshel, 1977): A-events are known only to the speaker (speaker authority) and B-events are known only to the addressee (addressee authority). Typical B-events include the addressee’s opinions, beliefs, bodily states, or professional expertise. The observation that authority to comment on events is unevenly distributed across speech-act participants is also explored in detail by Kamio (1997), who notes the infelicity of Japanese statements that target the addressee’s ‘territory of information’ unless these are marked by appropriate sentence-final particles, which serve to weaken the speaker’s epistemic claims and mitigate the force of such statements. Kamio’s conceptualisation of ‘territories of information’ is adopted by Heritage to define epistemic status as a relatively stable concept subject to socio-cultural conventions:

[W]e can consider relative epistemic access to a domain as stratified between actors such that they occupy different positions on an epistemic gradient (more knowledgeable [...] or less knowledgeable [...] which itself may vary in slope from shallow to deep ...). We will refer to this relative positioning as epistemic status, in which persons recognize one another to be more or less knowledgeable concerning some domain of knowledge[.] (Heritage, 2012b, p. 32)

The heuristic of an ‘epistemic gradient’ allows for a relative positioning of the speech-act participant’s knowledge-states and rights to knowledge. This notion has been used, for example, in cross-linguistic research on sentence-final particles that signal different kinds of questions (see Enfield, Brown, & de Ruiter, 2012; Hayano, 2012). The notion of epistemic gradient may be used to determine a speaker’s epistemic stance, as indicated by the speaker’s choice of sentence-type.

Heritage's efforts to detail how the epistemic statuses of speech participants shape turn-design enable us to look under the hood of the 'epistemic engine' of conversation (Heritage, 2012b). Indeed, language users are continuously keeping track of what others know and how their own knowledge can be related to the knowledge of others, and Heritage offers us a detailed and empirically grounded picture of how this 'epistemic ticker' works in everyday conversation.

There are, however, some issues that concern us in exploring the notion of 'engagement' from a cross-linguistic perspective, which are left mostly without comment in Heritage's work. One particularly important issue is what (linguistic) resources are available for conveying epistemic stance. While sentence-type has occupied a central role in research on English, linguistic forms signalling aspects of epistemic status and stance go well beyond sentence-type distinctions and may involve grammatical sub-systems that specifically target the perception, attention, and perspective of the speech participants, without requiring reformulation as interrogatives.

A final consideration is that Heritage's formulation of an epistemic gradient remains underspecified with respect to the individual commitments of the speech participants. That is, while a 'seesaw' gradient is conceptually useful, it veils the fact that the speaker's assumptions concerning the addressee's knowledge of some event are 'in the mind of the speaker' and do not necessarily correspond to the addressee's actual knowledge state (see below, Evans, 2006; cf. Bergqvist, 2015). The notion of multiple perspective, which we discuss in the next section, provides this underlying asymmetry with an explicit formulation, where the speech participant's points-of-view with respect to objects of discourse are calculated from the speaker's perspective.

4. Multiple perspective in grammar

As mentioned already, there is a clear asymmetry in the contrasts of epistemic distribution which engagement expresses. Whereas speakers have direct access to their own perspective, and can thus assert with confidence what they know, attend to, or perceive, in the case of the addressee they can only assume, to varying degrees of certainty. Assessments of the mental directedness of others therefore involve a type of complex perspective (Evans, 2006), which represents the speaker's assumption about the addressee's attentional state or access with respect to some state of affairs.¹⁹

[19] There have of course been long and thorny debates on how far recursive mutual inference about each other's mental states is possible: Sperber and Wilson (1986) argue that speaker and hearer must engage in pragmatic inference about each other at several recursive levels, and Scott-Phillips (2015) posits at least five levels of recursive mind-reading in any ostensive communicative act. For arguments that pragmatic inference is possible with a substantially less rich cognitive package than these scholars maintain, see Planer (2017a, 2017b).

As a caution that not all investigators have taken this as obvious, consider the discussion of definite articles in Givón (1989), and in particular his statement that definite descriptions are “inherently about knowledge by one mind of the knowledge of another mind” (p. 206). We do not share Givón’s epistemological optimism – that one mind can have knowledge of the knowledge of another mind. As a more accurate and epistemologically cautious characterisation, we prefer the formulation given in Hawkins (1978, p. 97): “the speaker when referring [and choosing between definite and indefinite articles – authors] must constantly take into consideration knowledge of various kinds which he assumes his hearer to have.”²⁰ This asymmetry – i.e., that assessments of knowledge or attention by the interlocutor are based on *assumptions* by the speaker – should be borne in mind throughout our discussion.

Multiple perspective constructions are constructions that “encode potentially distinct values, on a single semantic dimension, that reflect two or more distinct perspectives or points of reference” (Evans, 2006, p. 99). These are found in various parts of the grammar and fall into three kinds of perspectives: double, meta-, and complex perspective.

‘Double perspective’ is calculated with regard to two points of reference at once, each having equivalent epistemological status. An example is a demonstrative system like Japanese, where both the speaker’s and the addressee’s positions are taken into account when relating a figure to a location (e.g., Japanese: *kore* ‘speaker proximate’, *sore* ‘addressee proximate’, *are* ‘proximate to neither speaker nor addressee’; see Hinds, 1973). Double perspective constructions are likely to be limited to ‘transparent dimensions of experience’ such as space and time, as these do not require calculations regarding the attention and psychological state of others: the stated perspectival values of double perspective constructions are objectively verifiable. (As we shall see, however, this does not mean that spatial demonstratives cannot develop less epistemologically transparent uses, including psychological and attentional parameters – see §5, below.)

Meta- and complex perspective constructions are defined by the embedding of one perspective inside another. In meta-perspective constructions the perspective of one person is considered from the perspective of another. This can be seen in reported speech constructions such as, *He said (that) linguistics has high standards of evidence*, where the speaker asserts a report of another’s assertion, but does not directly represent the speaker’s position regarding the secondary assertion, i.e., *linguistics has high standards of evidence*.

[20] Arie Verhagen (p.c.) suggests a third position: that speakers can use the more optimistic common-ground scenario as a useful opening heuristic, at least in cases where there is mutually accessible evidence, then making adjustments (i.e., inferring asymmetries) when necessary – see also Verhagen (2015, especially section 3).

Complex perspective features the speaker's assertion of his/her own perspective along with that assumed by the speaker to hold for the addressee/other. The sentence *He is under the illusion that linguistics has high standards of evidence*, by using an anti-factive predicate in the main clause, simultaneously predicates one perspective of the embedded subject (who believes linguistics has high standards of evidence) and a different perspective of the speaker (who believes that any claim that linguistics has high standards of evidence is illusory). Summarising the contrast, a meta-perspective does not require the speaker's evaluation regarding the perspective of the other (although it may be present by implicature), whereas a complex perspective features non-defeasible assertions regarding both parties.

In the context of epistemic marking, multiple perspective constructions are arguably restricted to variants of meta- and complex perspective if one concedes that the perspective of the other necessarily is embedded in the speaker's perspective. The conceptualization of multiple perspective in epistemic marking targets the same issues that Heritage (2011, 2012a, 2012b, 2012c) details for epistemic status and stance, but with an increased focus on the different ways in which perspectives may be expressed, and what subsystems of language facilitate such expressions.

5. Demonstratives and the coordination of attention to objects and places

Es ist das Kernstück, es ist die bevorzugte Technik der anschaulichen Sprache, was wir als Zeigfeld beschreiben. (Bühler, 1934, p. 81)

What we describe as the deictic field is the core, the favoured technique of speech about perceptual things ... (Bühler, 1990, p. 95)

Arguably the most basic of intersubjective tasks in conversation is to coordinate the speaker's and addressee's attention on an object present in the context, by drawing the latter's attention towards that object through pointing or eye-gaze. After a long period when the typology of demonstrative systems was dominated by their spatial properties (Anderson & Keenan, 1985; Diessel, 1999a, 1999b; Dixon, 2003), the field is unveiling a growing number of cases where demonstratives can best be understood as grammatical devices for bringing one's interlocutor's attention into line with one's own (cf. Janssen, 2002). As Hausendorf (2003, pp. 257–9) puts it:

How can we account for the transition from single perceiving activities to mutually shared perception? ... Whenever sensory perception is to be extended or differentiated in order to make use of what can be seen,

heard, smelt or touched in the physical environment, deictic devices can be expected to make sure that these perceiving activities become mutually shared. ... I would propose to consider deixis as a device whose main function is to 'help' perceiving activities to become mutually shared communicative moves. ... Deixis allows visual perception to be perceived in itself.

Classic typologies of demonstrative systems (e.g., Anderson & Keenan, 1985) looked at the degrees of distance from the origo or speaker: two in (modern) English (*this/that*), three in Spanish (*este, ese, aquel*, using the analyses of Hottenroth, 1982, and Diessel, 1999a), and seven in Malagasy (but with an additional visible/invisible contrast that gives fourteen; Rasoloson & Rubino, 2005). These may then be elaborated by other spatial characteristics like up/down, upstream/downstream, etc. Despite their great variety, on these accounts all are fundamentally egocentric systems.

The next level of interpersonal complexity adds the possibility of taking other parties to the conversations as anchor point. Again, staying at the simplest level, entities can next be related to speaker, addressee, both, or neither, e.g., the three-way contrast in Japanese (*kore* speaker-proximal vs. *sore* addressee-proximal vs. *are* other), or the four-way contrast which is obtained in Quileute (Andrade, 1933, p. 252) by adding a fourth 'first inclusive' value: *xo'o* 'near the speaker', *so'o* 'near the second person', *sa'a* 'at a comparatively short distance from both', *á'tca'a* 'at a long distance'. Burarra (Glasgow & Glasgow, 1977) is similar, with some interesting further twists.²¹

Systems that take more than one conversational party as spatial anchor points may then be elaborated further by taking degrees of distance from two or more of these reference points. Abui, for instance (Kratochvil, 2007, 2011) has speaker-proximal, addressee-proximal, speaker-medial, addressee-medial, and distal (note that the speaker vs. addressee anchor point becomes irrelevant once the referent is far enough away), among other values bringing in factors like elevation. For example, one would say *do fala* for 'this house, near me', *to fala* for 'that house, near you', *o fala* or *lo fala* for 'that house, some distance

[21] In fact things are even more complex than this in Burarra, because there are proximal and distal forms for each of the four person-defined values, with the distal forms interacting with modes of evidence/knowledge/perceptual access depending on the person. Thus the first person inclusive distal form *-gata* is translated as 'that/those in sight or known to you and me', while the third person distal form *-gaba* is 'that/those out of sight there'. The second person proximal forms are still compatible with being close to the speaker as well, but imply either that they are habitually closer to the addressee, or near or known to him/her. For example, an out-of-town visitor to the regional capital, Darwin, on encountering locals there, might use the second person proximal form *ngunyunarda* because the addressees, who live there, would have greater knowledge of the current locale. This anticipates our discussion of asymmetries of knowledge in Part II, §2.

from me (but closer to me than you)', *yo fala* for 'that house, some distance from you (but closer to you than me)', and *oro fala* for 'that house (far from us both)'. Inuktitut (Denny, 1982) is another example of a language where there are two sets of demonstratives – speaker-anchored vs. other-anchored – where the second set may be anchored to a previous speaker, to the addressee, or to some other person or thing in the situation, which may not have been referred to before.

With these systems, we have now brought in *interpersonal* space – through the choice of speaker, addressee, both, or other as spatial anchor point – but not yet any intersubjective considerations, at least as far as most such systems are normally described – though one suspects that, for example, locations near the addressee are assumed to be more accessible to their attention, and even early accounts that focus on spatial semantics allow for metaphorical extensions into psychological domains.²²

At a third level of elaboration, perceptual modality enters the typology. We have already mentioned that Malagasy distinguishes visible from non-visible in addition to seven grades of distance. In Santali (Zide, 1972, digesting material from Bodding, 1929) demonstratives can add *-tɛ* for objects perceived visually and *-nɛ* for objects perceived by other senses which means, usually, aurally. Quileute (Andrade, 1933, p. 252), in addition to the four person-oriented forms mentioned above, has three forms for different types of partly or wholly invisible location: one for where they are nearby and maybe partly visible, one for where they are invisible but in a known location, and one where they are invisible and also in an unknown location.²³ The detailed analyses of the Yucatec Maya demonstrative system by Hanks (1990, 1999, 2007, 2009) show not only that there are formal contrasts based on a three-way contrast in sensory modality (visual, tactile, auditory/olfactory) in addition to distance, but also that the system is best understood as providing a “directive function ... whereby they direct an addressee to look, listen or take an object in hand” (Hanks, 1999, p. 124).

[22] For example, Anderson and Keenan (1985, p. 278) write that “spatial references [in deictic systems] serve as the basis, in most languages, for a variety of metaphorical extensions into other domains. ... notions such as ‘near to the speaker’ may be interpreted not only in the literal, physical sense, but also by extension to ‘psychological proximity’, i.e. vividness to the mind of the speaker”. They stop short, however, of mentioning more intersubjective metaphorisations such as we will see below.

[23] In Andrade’s words, the first of the invisible forms is used “when the location is near or when the speaker is in it, and hence, visible only in part”. And of the other two, he says “[t]heir use depends on whether the place is known to the speaker from previous direct experience, having been there, or whether he imagines the place or has heard of it” (1933, p. 252).

Our journey through demonstrative systems has thus led us into increasingly intersubjective terrain. Starting with a primarily spatial system,²⁴ we passed to systems which recognise other conversational participants as the anchor point for reckoning spatial relations, then on to those which direct the sensory modality which their interlocutors should use in searching for referents. We now raise the intercognitive status a final notch, examining demonstrative systems that explicitly encode the speaker's assumptions about whether the addressee has succeeded in locking onto the referent.

The first language for which this was shown clearly was Turkish, in studies by Aslı Özyürek (1998) and her colleagues Sotaro Kita (Özyürek & Kita, n.d.) and Aylin Küntay (Küntay & Özyürek, 2002, 2006). Turkish has a three-valued demonstrative system with three forms *bu*, *şu*, and *o*, which had previously been analysed as a person-based system on Japanese lines (e.g., Lyons, 1968) or as a distance-based system on Spanish lines (Bastuji, 1976; Serebrennikov & Gadzuyeva, 1979). However, these early analyses drew their base data from written texts in which the dynamics of face-to-face interaction could not be gauged accurately. Özyürek and her colleagues broke new ground by using videos of face-to-face interaction in which it was possible to track eye-gaze and pointing²⁵ behaviour at the same time as demonstrative use, leading to the following breakthrough.

Two of the Turkish demonstrative forms, *bu* and *o*, appear to be used roughly like English *this* and *that*, contrasting entities close to and distant from the speaker. It is the third form *şu* which is unusual compared to previously studied systems: it can be used for objects at any distance, but only if joint attention has not yet been established. This gives us the following set (Table 2), adjusting the first two for the fact that, unlike English, they require joint attention to be established in addition to specifying distance.

Consider the following example from the work of Özyürek and her colleagues. A teacher and two students are in a pottery class and one of the students wishes to refer to an object that is at the other end of the room. She points to it but the teacher's gaze has yet to fix on it (example (4) and Figure 1); at this point she uses the term *şu*:

- (4) *ya hocam şu oval mesela*
 well teacher NONMUTDEM oval for.example
 'well sir **that** oval(one) for example'

[24] To be clear here: we are not claiming that the systems considered so far disallow intercognitive readings (see footnote 22 on the 'metaphorical extensions' referred to by Anderson & Keenan, 1985), but rather that they contain no form whose meanings have been analysed as primarily intersubjective.

[25] Turkish speakers also use other means of indication, such as eyebrow-raising or raising the chin slightly (Göksel & Kerslake, 2005), though these were not mentioned in the Özyürek & Kita (n.d.) study.

TABLE 2. *The Turkish demonstrative system (after Özyürek & Kita, n.d.; Küntay & Özyürek, 2002)*

Form	Usage
<i>bu</i>	close to speaker, joint attention established
<i>o</i>	distant from speaker, joint attention established
<i>şu</i>	unspecified for distance, joint attention not yet established

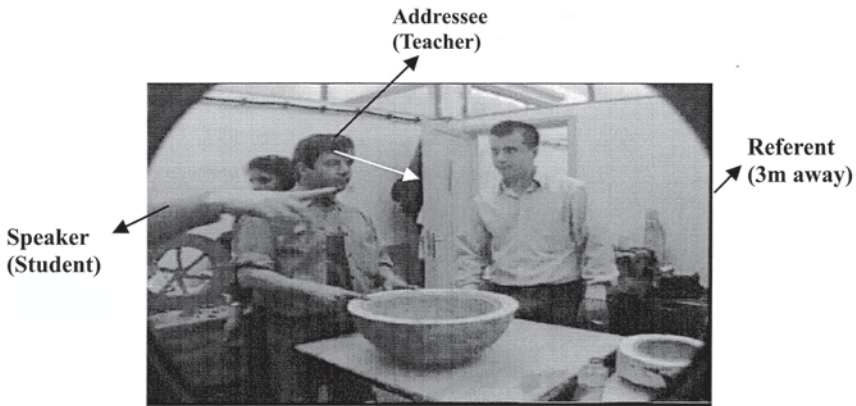


Fig. 1. Use of Turkish demonstratives (a).

In a second, more elaborated, utterance, in which she keeps pointing to the vase but the teacher's gaze has yet to lock onto it (example (5) and Figure 2), she continues to use the possessive form of *şu*, namely *şunun* 'of that one (which you have yet to identify)':

- (5) *şu-nun* *dış yüzey-in-e* *koy-up da*
 NONMUTDEM-GEN outer surface-GEN-DAT put-GER CONNec
 'by putting it on that thing's outer surface'

Finally the teacher's gaze moves up to follow the point and locate the referent (example (6) and Figure 3), and now the speaker switches to *o*, the form for distant but mutually attended objects (*o* is suffixed by (*n*)*dan* to mean 'from that'):

- (6) *ondan* *da olabilir*
 DIST:ABL and possible
 'That could be one as well.'

We can summarise how the Turkish deictic routine works in the following way: use a combination of pointing plus *şu* until you are sure of having achieved mutual attention on the object at issue, then proceed by using *bu* or *o* according to the distance to the referent.

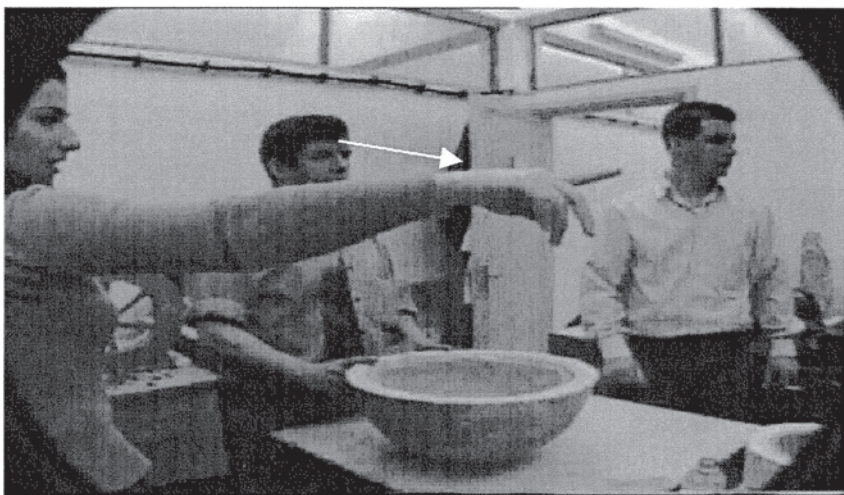


Fig. 2. Use of Turkish demonstratives (b).

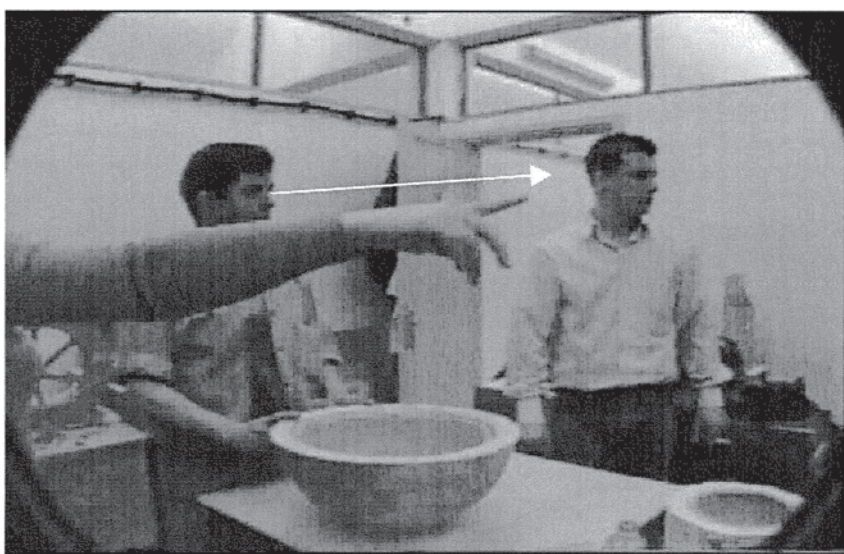


Fig. 3. Use of Turkish demonstratives (c).

Our second example comes from work by Niclas Burenhult (2003, 2008) on the Aslian language Jahai, spoken in Malaysia. Jahai has a set of eight demonstratives which can be arranged as in Table 3. The forms starting with a glottal stop (ʔ) are adverbials like ‘here’, while those starting with *t* are

TABLE 3. *Jahai demonstratives (Burenhult, 2003)*

Main parameter	Oppositions	Adverbial form	Nominal form	Referential characteristics in exophoric use
+ ACCESS	Speaker-anchored	<i>ʔəh</i>	<i>təh</i>	Accessible to speaker (proximal, perceptible, reachable, approachable, etc.)
	Addressee-anchored	<i>ʔon</i>	<i>ton</i>	Accessible to addressee (familiar, attended to)
- ACCESS	Speaker-anchored	<i>ʔaniʔ</i>	<i>taniʔ</i>	Inaccessible to speaker (distal, imperceptible, unreachable, inapproachable, etc.)
	Addressee-anchored	<i>ʔün</i>	<i>tün</i>	Inaccessible to addressee (unfamiliar, unattended to)
EXTERIOR	Speaker-anchored	<i>ʔadeh</i>	<i>tadeh</i>	Located outside speaker's side of speech perimeter
	Addressee-anchored	<i>ʔniʔ</i>	<i>tʔiʔ</i>	Located outside addressee's side of speech perimeter
ELEVATION	Superjacent	<i>ʔitih</i>	<i>titih</i>	Located above speech situation (overhead, uphill, or upstream)
	Subjacent	<i>ʔuyih</i>	<i>tuyih</i>	Located below speech situation (underneath, downhill, or downstream)

nominal demonstratives with meanings like ‘this’, but the logic of these two series is otherwise identical.

According to Burenhult, the Jahai conceive of conversation as a sort of container, and as “soon as a person addresses another person, they and the area between them become a connected spatial entity” (Burenhult, 2008, p. 116). The last four pairs in the table position objects with respect to that container. If we imagine it cut in half by a line between the speaker and the addressee, those on the speaker’s side but outside the container will be denoted by *tadeh*, those outside it but on the addressee’s side by *tʔiʔ*. Those conspicuously above or below the speech situation will be identified using the so-called superjacent or subjacent demonstratives from the ‘elevation’ set.

But it is the top four which interest us more here, and in particular the ‘addressee-anchored accessible’ *ton*. Burenhult obtained revealing data on this system using a ‘director-matching task’ where a ‘director’ has a photograph of different arrangements of objects, which he describes orally to a ‘matcher’ whose job is to reproduce the arrangement using real objects. In addition to his own photograph, the director can see the matcher and what he is setting out, whereas the matcher can only see his own objects and needs to rely on the director’s verbal description. Under these circumstances, discourses are produced which typically begin with the director’s introduction of a referent (e.g., ‘take the one which is flat and round’), proceed with a sequence of demonstrative exhortations by the

director as he monitors what the matcher is doing ('Underneath the one that has a hole. A different one, different one, different one. This one.') and end with a confirmation ('Yes, that one!'). The predominant pattern through these discourses is to culminate in the 'addressee-anchored accessible' *ton* after a series of other demonstratives giving spatial specification (examples (7) & (8)).

- (7) *tũn – tni? – ton* 'that (on your side but so far inaccessible to you) – that way over/on your side – that.one.now'
or
(8) *tami? – tami? – ton* 'this one (inaccess.) – this one (inaccess.) – that one now!'

The way the Jahai demonstratives track the speaker's monitoring of the addressee's attention is thus rather similar to Turkish, but the actual progression is almost the converse (see Table 4). The initial *şu* forms in Turkish give no spatial information of their own, merely telling the addressee to keep looking (in particular, to follow the point), but once lock-in has been achieved they give way to spatially specific forms (close to or far from speaker). In Jahai the forms used give much more spatial information as the progression unfolds – is it in the speaker's or the addressee's half of the container, or close to the speaker or the addressee? But once lock-in has been achieved, the form *ton* is used regardless of exact spatial position, as if the attentional accessibility of the object now makes spatial information unnecessary.

Before leaving these two systems, an observation is in order about the communicative ecology of pointing on the one hand and the demonstrative system on the other. The Turkish example makes it clear that achieving reference in conversation combines both gestural and linguistic elements as the demonstrative *şu* signals to the addressee to keep attending to the point. Indeed, Küntay and Özyürek (2002), who were puzzled by the fact that children still have not mastered the correct use of *şu* by the age of six despite the well-attested abilities of much younger children to monitor the gaze of adults, suggest that the delayed development is due to the extra cognitive demands of coordinating linguistic and gestural elements.²⁶

[26] Küntay and Özyürek (2002, p. 345) write that "These results might sound surprising in light of research indicating that joint attention is a very early communicative process that appears in infancy (Trevorthen 1998)". They go on to suggest that a "reason that we can propose is the integration of nonverbal factors with verbal expressions is a protracted developmental process (Goldin-Meadow, Alibali & Church 1993), and needs to develop further beyond 6 years of age. Especially when this integration is called for in a conversational task" (2002, p. 345). However, we believe that the late development of Turkish demonstrative use may in fact not be so surprising once we adopt a more graded view of how theory of mind develops, and note the fact that adult levels of theory of mind may not phase in till anywhere between five and eleven according to the specific test used (Saxe & Baron-Cohen, 2006). For a 'dual-process' theory of mind model that starts children off with an innate, rudimentary module available at birth, then refined through cultural learning at a much later age, see Apperly and Butterfill (2009).

TABLE 4. *Comparison of semantic contrasts during the search and lock-in phases in Turkish and Jahai demonstrative systems*

		Search phase	Lock-in phase	
Turkish	<i>şu</i>	'that which I am pointing to but which you haven't yet identified'	<i>bu</i>	'this which you have already identified'
			<i>o</i>	'that which you have already identified'
Jahai	<i>tūn</i>	'addressee-anchored unidentified'	<i>ton</i>	'addressee-anchored identified'
	<i>ʔni?</i>	'addressee-anchored exterior'		
	<i>tani?</i>	'speaker-anchored inaccessible'		
	<i>tāh</i>	'speaker-anchored accessible (to speaker)'		

On the other hand, in Jahai the use of actual pointing is much more limited. Within the experimental 'director-matcher' set-up, pointing was not an allowable part of the procedure. And in more naturalistic settings Burenhult mentions a number of reasons why pointing is much less common among Jahai than among most other cultures: communication often occurs while walking single-file along forest paths, or between spouses after dark, and in any case there are a number of cultural taboos against pointing. He goes on to suggest that the elaboration of the Jahai demonstrative system, which in effect gives a complex series of clues as to how the addressee should keep looking, compensates for the unavailability of pointing in many circumstances.

We draw our examination of demonstratives to a close by looking more briefly at two further examples where monitoring of the addressee's attention and expectations is relevant, though not in the sense we have seen of directly tracking whether they have latched onto the referent but rather in helping them assess its identification against previous expectations or searches.

The first comes from the Australian language Bininj Gun-wok, Gun-djeihmi dialect (Evans, 2003). Among a large number of demonstratives (and just giving the masculine forms, beginning with *na-*), an interesting part of this system is the intersection of distance with whether the speaker deems the addressee to have had some previous interest in the entity at issue. Let's say you are looking for something without success, and I spot it: I would then say either *nabernu* (if it is distant) or *nabehrnu* (the *h* represents a glottal stop) if it is close to hand. On the other hand, if I present something which I didn't think you had been interested in before (say I find a new plant which you didn't know existed) I could hold it up to you and say *nahni*. In other words, the system tracks pre-existing cognitive interest (or not) on the part of the addressee, and crosses this with distance.

A related phenomenon is attested for the Athapaskan language Kaska, namely the class of directionals (Moore, 2002, ch. 19; the term is also used by Golla, 1996), also referred to in the Athabaskanist literature as 'deictic/directionals'

(Rice, 1989), and ‘locationals’ (Henry & Henry, 1969). Leer (1989) has proposed that these derive from old sequences of a demonstrative plus a noun. Kaska directionals resemble demonstrative adverbs, and are built from two parts. The stem has spatial meanings like ‘off to the side’, ‘above’, ‘below’, ‘downstream’, ‘back down a trail’, or temporal meanings like ‘past’ or ‘future’. But it is the prefix which concerns us here, since these are sensitive to shared or unshared knowledge states.

Of crucial interest is the way three of the prefixes indicate different distributions of knowledge about the location across the speaker and addressee:

With reference to the more distant locations, the directional also indicates whether the speaker and the addressee know the exact location being referred to. For instance, the prefix *kùh-* is used when the *exact location is known by both the speaker and those they are addressing*. As other examples, the prefix *de-* is used when the *location is known by the speaker, but not those they are addressing*, and the prefix *ah-* is used when *neither the speaker nor their audience know the exact destination*, but only its approximate direction. (Moore, 2002, p. 404; italics added)

In terms of the four-way set of engagement values we found for Andoke (§2), this set covers three of the values: speaker-only, shared, and known to neither. It is only the fourth term – for the situation where the speaker does not know the exact location, but expects that the addressee might – that appears to be missing from this system.²⁷

Finally, we note that marking the mutual knowledge of speaker and addressee as regards an entity also appears to be relevant to what have been analysed as evidential morphemes either within or outside demonstrative systems, although these are generally less well understood and less documented cross-linguistically (see Jacques, in press). Storch and Coly (2014, p. 8) describe the suffix *-diyà* in Maaka (Nigeria) as indicating “that both speaker and hearer know or see the participant in question” (9). They further comment that this form originates from a Kanuri term meaning ‘surely, entirely, only’, highlighting the connection between joint witnessing and the establishment of truth (see also comments reproduced from Sillitoe, 2010, in Part II, §3).

- (9) *ʔaa-kè-díbb* *zùlùm-tò-diyà*
 COND-2SG:MASC-crush:PERV anus-POSS:3SG:FEM-JOINT:VIS
tà-kwáadà-ntí-minê *gè-ʔammà-à*
 3SG:FEM-throw:TR-ASSERT-OBJ:1PL LOC-water-DEF
 ‘If you crush her anus [that we can both see] she will definitely
 throw us into the water.’ (Storch & Coly, 2014, p. 197)

[27] Functionally, we might imagine that an interrogative form would fill this gap.

Across the world, in the South American language Lakondê (Telles & Wetzels, 2006) a nominal morpheme *-te-* ‘N. PROX’ is described as encoding both spatial distance and mutual visual perception. For example, ‘*sih-te-‘te* ‘house-N. PROX-REF’ is translated as ‘house which we see at a distance’.²⁸ Such nominal markers seem to be a genetic feature of Mamaindê languages and are especially elaborate in Southern Nambiquaran, which has aspect, tense, evidential, and engagement (termed ‘individual/collective verification’; Kroeker, 2001) marking on definite nouns (Lowe, 1999, p. 282). For example, the expression *wa3lin3su3ait3tã2* (numbers indicate tones) is glossed as ‘this manioc root that I, but not you, saw some time in the past’ and may be contrasted to *wa3lin3su3ait3ta3li2*, meaning ‘this manioc root that we (both) saw some time in the past’ (Lowe, 1999, p. 282; cf. Kroeker, 2001, pp. 45–6). The meaning contrast between individual and collective verification of the manioc root may be traced to the *-tã2* (individual verification) and the *-li2* (collective verification) suffix at the ends of the nominals. The complexity of Southern Nambiquaran, while staggering at first glance, is suggestive of the potential range of variation and the richness of such systems.

We have focused in such detail here on demonstratives because they are the syntactically simplest method of achieving mutual coordination – as investigators have pointed out, from Bühler²⁹ (as quoted in the epigraph to this section) on to Diessel:

demonstratives function to coordinate the interlocutors’ shared attentional focus. In the simplest case, the demonstrative is used to direct the addressee’s attention to a referent that previously was not in the shared attentional focus; in this case, the demonstrative creates a new joint focus of attention. However, demonstratives are also commonly used to direct the addressee’s attention from the current referent to a previously established referent or to differentiate between multiple referents that are already in the shared attentional focus. (Diessel, 2006, p. 470)

Demonstratives generally distinguish a reasonably large set of ontological categories – entities (this), places (here), times (now), manners (thus), and so forth, welded together with deictics into sets like *koko/soko/asoko* ‘here / there

[28] However, while both Maaka and Lakondê also have contrastive nominal markers that can indicate the visual perception of the speaker only, these do not form a clear paradigm with the mutual witness forms. In the Maaka case, the speaker-witness form *-mu* is only used with topicalised participants, while in Lakondê the speaker-only visual evidential *-ta-* does not encode spatial information and, unlike *-te-*, can be used on both nouns and verbs. It remains a possibility that joint ‘eye-witness’ markers are as much to do with mutual attentional status and affirmation as with information source and perception per se.

[29] And in fact this line of argument goes back through Steinthal (1891, p. 313) to Apollonius Dyscolus.

[by you] / there (away from us both)' in Japanese. However, the syntactic level at which they apply can be disarmingly simple. This makes it possible to use them in the most basic imaginable types of mini-dialogue, of the type discussed by Karcevski (1948/1969)³⁰ for Russian pairs like *Ty kuda? Tuda.* ('You ('re going) whither?' 'Thither (accompanied by a suitable gesture).'); see Diessel (2003) and Evans (2012) for further discussion of these 'dialogic parallelisms'.

These Karcevskian dialogues are possible because the semantics of the deictic expressions is essentially self-contained:³¹ a pairing of a deictic value (e.g., proximal vs. distal) and an ontological one (e.g., place, or time, or manner). In Part II of this paper, we will pass to a number of systems where attentional coordination has been expanded to the point where it concerns not just objects, but the broader domain of events and the epistemic background to talking about them. There are some important differences between engagement as it can apply to objects (especially objects that are present in the speech situation) and as it applies to events and situations, which may require increased abstraction in reference, and, once in the past, are not available for ostension and must rather be remembered, learnt, believed, etc. We explore the complexity of encoding the differential accessibility of events using data from languages of the Americas, Papua New Guinea, and Northern India (for example: Did the speaker directly experience this event? Did the addressee experience it, too?). Finally, we see that, as regards the category of engagement, the distinction between objects and states of affairs is not so hard-and-fast: Abui shows that a diachronic pathway between the two can be traced via the increased functionality of demonstrative forms. And so we move from the world of entities, as discussed in Part I of this study, to the world of events, the topic of Part II.

Acknowledgments

We thank Laura Michaelis-Cummings for the opportunity to put forward the ideas in this paper through an invited paper to *Language and Cognition*. For institutional support of the research underlying it, we are grateful to the Australian Research Council (grants DP0878126 'Language and Social Cognition: the Design Resources of Grammatical Diversity' and

[30] "Les interrogatifs indiquent le 'vide' dans le tissu sémantique, c'est-à-dire ce qui est ignoré. ... La réponse la plus concrète à la question *Ty kuda?* 'Où vas-tu?' serait naturellement *Tuda* 'Là-bas' accompagnée du geste correspondant. C'est pourquoi la formule du dialogue d'information est *k/t*, c'est-à-dire 'ignoratif' ≈ déictique."

[31] Though of course in many languages demonstratives can carry information about broader syntactic context, such as gender, number, or case. Nonetheless, in the most basic situations of the type Question: Demonstrative, case is not relevant, and gender and number can generally be determined from the situation.

FL130100111 ‘The Wellsprings of Linguistic Diversity’), the Australian Research Council Centre of Excellence for the Dynamics of Language, the Alexander von Humboldt Foundation (Anneliese Maier Forschungspreis to Evans), the Swedish Research Council (dnr. 2011-2274), and the Netherlands Organisation for Scientific Research NWO (Netherlands Organisation for Scientific Research), Veni award 275-89-024, ‘Learning the senses: Perception verbs in child–caregiver interaction’, as well as to our respective host institutions: the Australian National University, Stockholm University, and Radboud Universiteit in Nijmegen. The ideas in this paper have emerged from discussions with many people, and we particularly thank the following: Niclas Burenhult, Bill Hanks, Sotaro Kita, Jon Landaburu, and Aslı Özyürek; we additionally thank Sotaro Kita and Aslı Özyürek for permission to reproduce figures from an unpublished paper they wrote on Turkish demonstratives that has influenced us deeply. Ron Planer, Matt Spike, Alan Rumsey, and Arie Verhagen gave much-appreciated helpful critical comments on an earlier version of this manuscript, as did two anonymous referees, and Susan Ford did an immaculate job in checking, formatting, and editing it.

REFERENCES

- Aksu-Koç, A., & Slobin, D. (1986). A psychological account of the development and use of evidentials in Turkish. In W. Chafe & J. Nichols (Eds.), *Evidentiality: the linguistic coding of epistemology* (pp. 159–167). Norwood, NJ: Ablex.
- Anderson, S., & Keenan, E. L. (1985). Deixis. In T. Shopen (Ed.), *Language typology and linguistic description, Vol. 3, Grammatical categories and the lexicon* (pp. 259–308). Cambridge: Cambridge University Press.
- Andrade, M. J. (1933). Quileute. In F. Boas (Ed.), *Handbook of American Indian languages 3* (pp. 151–292). New York: Columbia University Press.
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, **116**(4), 953–970.
- Austin, J. L. (1962). *How to do things with words*. Oxford: Clarendon Press.
- Bastuji, J. (1976). *Les relations spatiales en turc contemporain*. Paris: Klincksieck.
- Bergqvist, H. (2015). Epistemic marking and multiple perspective: an introduction. *Sprachtypologie und Universalienforschung (STUF)*, **68**(2), 1–19.
- Biber, D., & Finegan, E. (1989). Styles of stance in English: lexical and grammatical marking of evidentiality and affect. *Text* **9**(1), 93–124.
- Bodding, P. O. (1929). *A Santali dictionary*. Oslo: J. Dybwad.
- Brown, D., Chumakina, M., & Corbett, G. (Eds.) (2013). *Canonical morphology and syntax*. Oxford: Oxford University Press.
- Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: G. Fischer.
- Bühler, K. (1990). *Theory of language: the representational function of language*, trans. D. F. Goodwin. Amsterdam: Benjamins.
- Burenhult, N. (2003). Attention, accessibility, and the addressee: the case of the Jahai demonstrative *ton*. *Pragmatics*, **13**, 363–379.
- Burenhult, N. (2008). Spatial coordinate systems in demonstrative meaning. *Linguistic Typology*, **12**(1), 99–142.
- Clark, H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge: Cambridge University Press.

- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Science*, **13**, 148–153.
- Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Science*, **366**(1567), 1149–1157.
- Denny, J. P. (1982). Semantics of the Inuktitut (Eskimo) spatial deictics. *International Journal of American Linguistics*, **48**(4), 359–384.
- Desclés, J.-P. (2009). Prise en charge, engagement et désengagement. *Langue Française*, **162**, 29–54.
- Diessel, H. (1999a). *Demonstratives: form, function, and grammaticalization* (Typological Studies in Language 42). Amsterdam: Benjamins.
- Diessel, H. (1999b). The morphosyntax of demonstratives in synchrony and diachrony. *Linguistic Typology*, **3**, 1–49.
- Diessel, H. (2003). The relationship between demonstratives and interrogatives. *Studies in Language*, **27**(3), 635–655.
- Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, **17**(4), 463–489.
- Dixon, R. (2003). Demonstratives. *Studies in Language*, **27**(1), 61–112.
- Du Bois, J. (2007). The stance triangle. In R. Engelbretson (Ed.), *Stancetaking in discourse* (pp. 139–182). Amsterdam: Benjamins.
- Enfield, N., Brown, P., & de Ruiter, J. (2012). Epistemic dimensions of polar questions: sentence final particles in comparative perspective. In J. de Ruiter (Ed.), *Questions: formal, functional and interactional perspectives* (pp. 193–221). Cambridge: Cambridge University Press.
- Enfield, N., & Levinson, S. (2006). *Roots of human sociality*. Oxford: Berg.
- Engelbretson, R. (Ed.) (2007). *Stancetaking in discourse: subjectivity, evaluation, interaction*. Amsterdam: Benjamins.
- Evans, N. (2003). *Bininj Gun-wok: a pan-dialectal grammar of Mayali, Kunwinjku and Kune*. Canberra: Pacific Linguistics.
- Evans, N. (2006). View with a view: towards a typology of multiple perspective. *Berkeley Linguistics Society (BLS)*, **32**, 93–120.
- Evans, N. (2012). Nen assentives and the problem of dyadic parallelisms. In A. C. Schalley (Ed.), *Practical theories and empirical practice: facets of a complex interaction* (pp. 159–183). Amsterdam: Benjamins.
- Givón, T. (1989). *Mind, code and context: essays in pragmatics*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Glasgow, D., & Glasgow, K. (1977). *Burrara work papers – texts*. Darwin: Summer Institute of Linguistics.
- Goffman, E. (1974). *Frame analysis: an essay on the organization of experience*. London: Harper and Row.
- Goffman, E. (1981). *Forms of talk*. Philadelphia, PA: University of Pennsylvania Press.
- Göksel, A., & Kerslake, C. (2005). *Turkish: a comprehensive grammar*. New York: Routledge.
- Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transition in concept acquisition: using the hand to read the mind. *Psychological Review*, **100**, 279–297.
- Golla, V. (1996). Sketch of Hupa, an Athapaskan language. In I. Goddard (Ed.), *Handbook of North American Indians, Vol. 17, Languages* (pp. 364–389). Washington, DC: Smithsonian Institution.
- Goody, E. (Ed.) (1995). *Social intelligence and interaction: expressions and implications of the social bias in human intelligence*. Cambridge: Cambridge University Press.
- Guentchéva, Z. (2011). L'opération de prise en charge et la notion de médiativité. In P. Dendale & D. Coltier (Eds.), *La prise en charge énonciative: Etudes théoriques et empiriques* (Champs Linguistiques) (pp. 117–142). Brussels: De Boeck-Duculot.
- Guentchéva, Z., & Landaburu, J. (2007). Introduction. In Z. Guentchéva & J. Landaburu (Eds.), *L'énonciation médiatisée II: Le traitement épistémologique de l'information: Illustrations amérindiennes et caucasiennes* (pp. 1–19). Leuven: Peeters.
- Gundel, J., & Fretheim, T. (2006). Topic and focus. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 175–196). Oxford: Blackwell.
- Hanks, W. (1990). *Referential practice, language and lived space among the Maya*. Chicago, IL: University of Chicago Press.

- Hanks, W. (1999). Indexicality. *Journal of Linguistic Anthropology*, 9(1), 124–126.
- Hanks, W. (2007). Person reference in Yucatec Maya conversation. In N. J. Enfield & T. Stivers (Eds.), *Person reference in interaction: linguistic, cultural, and social perspectives* (pp. 149–171). Cambridge: Cambridge University Press.
- Hanks, W. (2009). Fieldwork on deixis. *Journal of Pragmatics*, 41(1), 10–24.
- Hausendorf, H. (2003). Deixis and speech situation revisited: the mechanism of *perceived perception*. In F. Lenz (Ed.), *Deictic conceptualisation of space, time and person* (pp. 249–269). Amsterdam: Benjamins.
- Hawkins, J. (1978). *Definiteness and indefiniteness: a study in reference and grammaticality prediction*. London: Croom Helm.
- Hayano, K. (2011). Claiming epistemic primacy: *yo*-marked assessments in Japanese. In T. Stivers, L. Mondada, & J. Steensig (Eds.), *The morality of knowledge in conversation* (pp. 58–81). Cambridge: Cambridge University Press.
- Hayano, K. (2012). *Territories of knowledge in Japanese interaction*. Unpublished doctoral dissertation, Radboud University.
- Henry, D., & Henry, K. (1969). Koyukon locations. *Anthropological Linguistics*, 11, 136–142.
- Heritage, J. (2002). Oh-prefaced responses to assessments: a method of modifying agreement/disagreement. In C. Ford, B. Fox, & S. Thompson (Eds.), *The language of turn and sequence* (pp. 196–224). New York: Oxford University Press.
- Heritage, J. (2011). Territories of experience, territories of knowledge: empathic moments in interaction. In T. Stivers, L. Mondada, & J. Steensig (Eds.), *The morality of knowledge in conversation* (pp. 159–183). Cambridge: Cambridge University Press.
- Heritage, J. (2012a). Epistemics in action: action formation and territories of knowledge. *Research on Language & Social Interaction*, 45(1), 1–29.
- Heritage, J. (2012b). The epistemic engine: sequence organization and territories of knowledge. *Research on Language and Social Interaction*, 45(1), 30–52.
- Heritage, J. (2012c). Beyond and behind the words: some reactions to my commentators. *Research on Language and Social Interaction*, 45(1), 76–81.
- Heritage, J. (2013). Epistemics in conversation. In J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 370–394). Malden, MA: Wiley-Blackwell.
- Heritage, J., & Raymond, G. (2005). The terms of agreement: indexing epistemic authority and subordination in assessment sequences. *Social Psychology Quarterly*, 68(1), 15–38.
- Heritage, J., & Raymond, G. (2012). Navigating epistemic landscapes: acquiescence, agency and resistance in responses to polar questions. In J.-P. De Ruiter (Ed.), *Questions: formal, functional and interactional perspectives* (pp. 179–192). Cambridge: Cambridge University Press.
- Hinds, J. (1973). Some remarks on *soo su-*. *Papers in Japanese Linguistics*, 2, 18–30.
- Hottenroth, P.-M. (1982). The system of local deixis in Spanish. In J. Weissenborn & W. Klein (Eds.), *Here and there: cross-linguistic studies on deixis and demonstration* (pp. 133–154). Amsterdam: Benjamins.
- Hylland, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*, 7(2), 173–192.
- Jacques, G. (in press). Non-propositional evidentiality. In Aikhenvald, A. (Ed.), *The Oxford handbook of evidentiality*. Oxford: Oxford University Press.
- Jakobson, R. (1990 [1957]). Shifters and verbal categories. In L. Waugh & M. Monville-Burston (Eds.), *On language* (pp. 386–392). Cambridge, MA: Harvard University Press.
- Janssen, T. (2002). Deictic principles of pronominals, demonstratives and tenses. In F. Brisard (Ed.), *Grounding: the epistemic footing of deixis and reference* (pp. 151–193). Berlin: Mouton de Gruyter.
- Jespersen, O. (1924). *The philosophy of grammar*. London: Routledge.
- Kamio, A. (1997). *Territory of information* (Pragmatics and Beyond New Series 48). Amsterdam: Benjamins.
- Karcevski, S. (1948). Sur la parataxe et la syntaxe en russe. *Cahiers Ferdinand de Saussure* 7, 33–38. (Reprinted in Robert Godel (Ed.), 1969, *A Geneva School reader in linguistics* (pp. 212–227), Bloomington, IN: Indiana University Press.)

- Kirsner, R. S. (2003). On the interaction of the Dutch pragmatic particles *hoor* and *hè* with the Imperative and Infinitivus Pro Imperativo. In A. Verhagen & J. van de Weijer (Eds.), *Usage-based approaches to Dutch* (pp. 59–96). Utrecht: LOT.
- Kockelman, P. (2004). Stance and Subjectivity. *Journal of Linguistic Anthropology*, **14**, 127–150.
- Kratochvil, F. (2007). *A grammar of Abui: a Papuan language of Alor*. Utrecht: LOT.
- Kratochvil, F. (2011). *Demonstratives as markers of stance: evidence from Abui*. Unpublished manuscript.
- Kroeker, M. (2001). A descriptive grammar of Nambikuara. *International Journal of American Linguistics*, **67**(1), 1–87.
- Küntay, A., & Özyürek, A. (2002). Joint attention and the development of the use of demonstrative pronouns in Turkish. In B. Skarabela, S. Fish, & A. H.-J. Do (Eds.), *Proceedings of the 26th Boston University Conference on Language Development (BUCLD 26)* (pp. 336–347). Somerville, MA: Cascadilla Press.
- Küntay, A., & Özyürek, A. (2006). Learning to use demonstratives in conversation: What do language specific strategies in Turkish reveal? *Journal of Child Language*, **33**, 303–320.
- Labov, W., & Fanshel, D. (1977). *Therapeutic discourse: psychotherapy as conversation*. New York: Academic Press.
- Lambrecht, K. (1994). *Information structure and sentence form: topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Landaburu, J. (2005). Expresión gramatical de lo Epistémico en Algunas Lenguas del Norte de Suramerica. In *Proceedings of the conference on indigenous languages of Latin American II, University of Texas at Austin, 27th–29th October 2005*.
- Landaburu, J. (2007). La modalisation du savoir en langue andoke (Amazonie colombienne). In Z. Guentchéva & J. Landaburu (Eds.), *L'énonciation médiatisée II: Le traitement épistémologique de l'information; Illustrations amérindiennes et caucasiennes* (pp. 23–47). Leuven: Peeters.
- Leer, J. (1989). Directional systems in Athapaskan and Na-Dene. In E.-D. Cook & K. Rice (Eds.), *Athapaskan linguistics: current perspectives on a language family* (pp. 575–619). New York: Mouton de Gruyter.
- Levinson, S. (1979). Activity types and language. *Linguistics*, **17**, 365–399.
- Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Lowe, I. (1999). Nambiquara. In R. Dixon & A. Aikhenvald (Eds.), *Amazonian languages* (pp. 268–291). Cambridge: Cambridge University Press.
- Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
- McCawley, J. D. (1981). Notes on the English perfect. *Australian Journal of Linguistics*, **1**, 81–90.
- McCoard, R. W. (1978). *The English perfect: tense-choice and pragmatic inferences*. Amsterdam: North-Holland.
- Michael, L. (2012). Nanti self quotation: implications for the pragmatics of reported speech and evidentiality. *Pragmatics and Society*, **3**(2), 321–357.
- Michaelis, L. A. (1994). The ambiguity of the English present perfect. *Journal of Linguistics*, **30**, 111–157.
- Moore, P. (2002). *Point of view in Kaska historical narratives*. Unpublished doctoral dissertation, University of Indiana.
- Moore, R. (2015). Meaning and ostension in great ape gestural communication. *Animal Cognition*, **19**(1). Online: <doi:10.1007/s10071-015-0905-x>.
- Nuyts, J. (2006). Modality: overview and linguistic issues. In W. Frawley (Ed.), *The expression of modality* (pp. 1–26). Berlin: Mouton de Gruyter.
- Özyürek, A. (1998). An analysis of the basic meaning of Turkish demonstratives in face-to-face conversational interaction. In S. Santi, I. Guaitella, C. Cave, & G. Konopczynski (Eds.), *Oralite et gestualite: Communication multimodale, interaction; actes du colloque ORAGE 98* (pp. 609–614). Paris: L'Harmattan.
- Özyürek, A., & Kita, S. (n.d.). *Joint attention and distance in the semantics of Turkish and Japanese demonstrative systems*. Unpublished manuscript.

- Planer, R. (2017a). Protolanguage might have evolved before ostensive communication. *Biological Theory*, **12**, 72–84.
- Planer, R. (2017b). Talking about tools: Did early pleistocene hominins have a protolanguage? *Biological Theory*, online: <doi:10.1007/s13752-017-0279-1>.
- Prince, E. (1981). Toward a taxonomy of given–new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255). New York: Academic Press.
- Rasoloson, J., & Rubino, C. (2005). Malagasy. In K. Adelaar & N. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 456–488). London: Routledge.
- Rice, K. (1989). *A grammar of Slave*. Berlin: Mouton de Gruyter.
- Sacks, H. (1987). On the preferences for agreement and contiguity in sequences in conversation. In G. Button & J. Lee (Eds.), *Talk and social organization* (pp. 54–69). Clevedon: Multilingual Matters.
- Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, **50**, 696–735.
- Saxe, R., & Baron-Cohen, S. (Eds.) (2006). *Theory of mind: a special issue of Social Neuroscience*. London: Psychology Press.
- Schegloff, E. (2007). *Sequence organization in interaction: a primer in conversation analysis, Vol. 1*. Cambridge: Cambridge University Press.
- Scott-Phillips, T. (2015). *Speaking our minds: why human communication is different and how language evolved to make it special*. London: Palgrave Macmillan.
- Searle, J. (1969). *Speech acts: an essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Serebrennikov, B., & Gadzuyeva, N. (1979). *Sravnitel'no-istoricheskaya grammatika tyurkskikh yazykov*. Baku: Izdatel'stvo "Maarif"
- Sillitoe, P. (2010). Trust in development: some implications of knowing in indigenous knowledge. *Journal of the Royal Anthropological Institute* (N.S.), **16**, 12–30.
- Slobin, D., and Aksu-Koç, A. (1982). Tense, aspect and modality in the use of the Turkish evidential. In P. Hopper (Ed.), *Tense-Aspect: between semantics and pragmatics* (pp. 185–200). Philadelphia, PA: Benjamins.
- Sperber, D., & Wilson, D. (1986). *Relevance: communication and cognition*. Oxford: Blackwell.
- Steinthal, H. (1891). *Geschichte der Sprachwissenschaft bei den Griechen und Römern mit besonderer Rücksicht auf die Logik*, 2nd ed. Berlin: Ferd. Dümmler.
- Stivers, T., & Rossano, F. (2010). Mobilizing response. *Research on Language in Social Interaction*, **43**(1), 1–31.
- Storch, A., & Coly, J. (2014). The grammar of knowledge in Maaka (Western Chadic, Nigeria). In A. Aikhenvald & R. Dixon (Eds.), *The grammar of knowledge: a cross-linguistic typology* (pp. 190–208). Oxford: Oxford University Press.
- Telles, S., & Wetzels, L. (2006). Evidentiality and epistemic mood in Lakondé. In E. Carlin & G. Rowicka (Eds.), *What's in a verb? Studies of the verbal morphology of the languages of the Americas* (pp. 235–252). Utrecht: LOT.
- Timberlake, A. (2007). Modality: overview and linguistic issues. In T. Shopen (Ed.), *Language typology and syntactic description*, 2nd ed., 3 Vols. (pp. 280–333). Cambridge: Cambridge University Press.
- Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: Bradford Books.
- Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, MA: Harvard University Press.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral & Brain Sciences*, **28**, 675–735.
- Tournadre, Nicolas, & LaPolla, R. J. (2014). Towards a new approach to evidentiality: issues and directions for research. *Linguistics of the Tibeto-Burman Area*, **37**(2), 240–263.
- Trevarthen, C. (1979). Communication and cooperation in early infancy: a description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech* (pp. 321–347). Cambridge: Cambridge University Press.
- Trevarthen, C. (1998). The concept and foundations of infant intersubjectivity. In S. Braten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 15–46). New York: Cambridge University Press.

- Verhagen, A. (2005). *Constructions of intersubjectivity: discourse, syntax and cognition*. Oxford: Oxford University Press.
- Verhagen, A. (2015). Grammar and cooperative communication. In E. Dąbrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (Handbooks of Linguistics and Communication Science 39) (pp. 232–252). Berlin: De Gruyter Mouton.
- Wilkins, D. (1986). Particle/clitics for criticism and complaint in Mparntwe Arrernte (Aranda). *Journal of Pragmatics*, **10**(5), 575–596.
- Wu, R.-J. R. (2004). *Stance in talk: a conversation analysis of Mandarin final particles*. Amsterdam: Benjamins.
- Wu, R.-J., & Heritage, J. (2017). Particles and epistemics: convergences and divergences between English and Mandarin. In Lerner, G. & Raymond, G. (Eds.), *Enabling human conduct: naturalistic studies of talk-in-interaction in honor of Emanuel A. Schegloff* (pp. 273–298). Amsterdam: Benjamins.
- Zide, N. (1972). A Munda demonstrative system: Santali. In J. Barrau, J. Thomas, L. Bernot, & A. Haudricourt (Eds.), *Langues et techniques, nature et société* (pp. 267–274). Paris: Klincksieck.