CHAPTER 19

Non-verbal communication and context: multi-modality in interaction¹
Pauline Madella and Tim Wharton

1. Introduction

For obvious reasons, the study of linguistics has focussed on verbal modality. Phonologists study the sound structures of language, syntacticians the grammatical structures, semanticists the meaning inherent in language. There is, however, a bigger picture, and those of us working in the sub-discipline of linguistic pragmatics find ourselves quite regularly faced with a dilemma. In contrast to semantics, which studies context-independent sentences, the aim of a pragmatic theory is to explain how utterances are understood, and utterances, of course, incorporate both verbal and non-verbal modalities. We do not deliver our words in a monotonic, robot-like voice. We impose quasi-musical contours on the words we say and move our faces and hands – in sometimes highly idiosyncratic ways – in order to guide the hearer to our intended meanings. Sometimes, such behaviours tell us more about the emotional state of a friend or colleague than the words they utter. Whether we are researchers, or simply trying to help a friend, non-verbal modalities cannot be ignored.

This paper is an introduction to the major role non-verbal modes play in communication and, in so doing, promotes a ‘multimodal’ view of human communication and looks at how the different modes identified not only constrain the construction of context but effectively guide our interlocutors to our intended meaning. We begin by addressing two questions:

(1) What is multimodality?
(2) How do non-verbal behaviours communicate?

In the next section we address question (1) and explore the relationship between the terms ‘paralanguage’, ‘multimodality’ and ‘non-verbal’. In section three we introduce the different categories of non-verbal behaviour and sketch an answer to (2) from the perspective of relevance theory (Wilson and Wharton, 2006; Wharton, 2009). In section four we address a third and final question:

(3) How can we move towards a principled pragmatic account of multimodality in interaction?

We argue that a pragmatic account based on relevance theory can be carried over to a wide range of multimodal phenomena, requires none of the intermediate levels of description that necessarily form part of less pragmatic accounts and falls out naturally from general communicative principles.

¹ Tim Wharton would like to thank Mengyang Qiu and Chara Vlachaki for help preparing this manuscript.
2. What is multimodality?

The argument that non-verbal modalities play a major role in accounts of human communication is not new. Insights from early pragmatics into the significance of paralinguistic behaviours in interaction certainly gives us a taste of the non-verbal dominance view of communication. Whilst Birdwhistell (1970, p. 158) reports that ‘probably no more than 30 to 35 percent of the social meaning of a conversation or an interaction is carried by the words’, Mehrabian’s 7%-38%-55% rule (1971) stipulates that 55 per cent of the information that we convey in a given exchange is communicated via our body language, 38 per cent via our tone of voice or how we say what we say and seven per cent from what is said (i.e. the words themselves). The importance of acknowledging the role of paralinguistic phenomena in face-to-face interaction is articulated by Abercrombie (1968, p. 65): ‘The conversational use of spoken language cannot be properly understood unless paralinguistic elements are taken into account.’ While Stevick (1982, p. 163) describes non-verbal communication as providing ‘the surface on which the words are written and against which they must be interpreted’. In this paper, we view multimodality as close to Claudel’s notion of L’oeil écoute (1946; Madella, 2021), encompassing both visible and audible modes of communication and interpretation, which are to ‘converge to contribute to a final meaning or message of which they are an intrinsic part.’ (Riley, 1979, p. 143). 'Multimodal’ inputs – what Adam Kendon refers to as ‘poly-modalic’ (2014a, p. 67) – invite the eyes to be as alert as the ears. Linguistic communication is, beyond any reasonable doubt, multimodal. Pragmatic accounts of linguistic communication must be multimodal also.

The above accounts coincided with the development of the term ‘paralanguage’ as first introduced by Trager in 1958. The term was originally introduced to refer to the way something is said, rather than what is said (Trager, 1958). In Paralanguage: a first approximation (1958), Trager described paralanguage as centred around modifications of the voice and other vocal features, to the exclusion of gestural ones. As Poyatos puts it (1983, p. 129):

What truly gives the spoken words their total meaning (which at any rate is not contained only in them) are a series of vocal/narial voice modifications and independent sounds and meaningful silences which today we subsume under paralanguage; and that, if visually perceived, those verbal expressions are accompanied by a great number of facial, manual and bodily gestures, gaze activities, manners, postures, postural shifts and stills, which constitute kinesics.

At the same time, the study of kinesics was beginning to take on a life of its own and that term came to be used to refer to all bodily movements, including those that are not intentionally communicative.

A subtly different stance was then proposed by Abercrombie (1968, p. 68), who rejected what he described as the ‘unfortunate separation of the visible and the audible components’. Abercrombie integrated both audible and visible paralinguistic elements into his definition of paralinguistic phenomena. However, he made a point of distinguishing elements that are conversational in nature from those that are not, and only those that are intentionally
communicative or ‘consciously controllable’ qualify as paralinguistic according to his definition (Abercrombie, 1968, p. 65).

In this chapter, we adopt Abercrombie’s inclusive definition of paralinguistic behaviour, on the grounds that the pragmatics of those paralinguistic modes of communication is to be found in the way that they interact with each other. We equate paralanguage with non-verbal or multimodal communication. However, it is worth noting immediately that Abercrombie’s exclusion of spontaneous, uncontrollable behaviours is controversial. According to Wharton (2009) spontaneous behaviours might be deliberately and overtly shown in a way that makes it in a sense ‘intentional’. Such behaviours might also be shown whilst giving my interlocutor the impression that it was not done intentionally. We touch on these later.

Paralanguage functions to constrain the context and guide the hearer to the speaker’s intended meaning. This is precisely what makes these inputs ‘multimodal’ or, as Kendon (2014a, p. 67) terms them, ‘poly-modalic’. The parallels between all these inputs are so strong that a unified, homogeneous account of para-/non-linguistic behaviours seems to be required, one that embraces the fact that they are, for the most part, closely interlinked. Indeed, an inclusive definition of paralinguistic phenomena is precisely what allows us to argue for an inferential model of (verbal and non-verbal) communication from the perspective of relevance theory. In the next section we turn to prosody, one element of paralanguage that falls under Abercrombie’s broad definition, and one that has received a great deal of attention in the pragmatics literature.

3. The relevance of prosody: how do non-verbal behaviours communicate?

When it comes to prosody, one point upon which there is broad agreement in the literature is that it comes in a range of different flavours. Some aspects of prosody feel entirely natural, in the sense that the link between form and meaning is far from arbitrary (or, to use Grice’s 1957 term, non-natural). For example, a person speaking to a young child will typically adopt a higher pitch, which I perceived to be non-threatening²; when people are angry, the volume of their voice tends to increase. Others are clearly language-specific (Gussenhoven, 2002). However, drawing a line between the natural and the non-natural is notoriously difficult. According to Bolinger (1964: 282) prosody exists, at best, at ‘the edge of language’, and observes that even those fully non-natural elements of prosody elements betray a degree of naturalness.

Early work in prosody did not refer much to ideas in pragmatics, largely because pragmatic theories were perceived as underdeveloped. So while historically there is a huge phonological literature providing detailed analyses of English prosody and how it relates to meaning (Halliday, 1963, 1967; O’Connor and Arnold, 1973; Brazil, 1975; Ladd, 1978, 1996; Bolinger, 1983ab; Ward and Hirschberg, 1988; Hirschberg and Ward, 1995; Pierrehumbert and Hirschberg, 1990; Gussenhoven, 1984, 2002, 2006; Chen and Gussenhoven, 2003), little of this work took place within a clearly articulated theoretical pragmatic framework. However, since the 1990s, there has been an increase in work on the semantics and pragmatics of prosody thanks largely to advances in how pragmatic inference is understood (Clark, 2007; Clark and

² This is a fairly universal phenomenon, presumably based on the fact that a higher-pitched voice equals a shorter larynx, equals a smaller creature, equals less danger.
Recent work in psychology and cognitive science has built on insights gained in the 1940s and 1950s on ‘bounded rationality’, the idea that an individual’s thought processes are limited by both the finite nature of cognition and the time available to make decisions. It has been proposed that evolution has left humans with economical rules-of-thumb, which enable us to make the most of our finite cognitive capacity: ‘simple, intelligent heuristics capable of making near optimal inferences’ (Kahneman, 2011, p. 168). Work within cognitively oriented approaches to pragmatics has allowed us to suggest new ways of thinking about prosody, and new ways of framing original hypotheses as to how it works.

According to relevance theory, cognition is geared towards the maximisation of relevance, where information is relevant to the extent that it provides communicative effects for relatively minimal processing effort. Other things being equal, the more communicative effects gained, and the less processing effort expended in gaining those effects, the greater the relevance of the information to the individual who processes it. This linguistic theory chimes harmoniously with theories developed in cognitive science. Indeed, cognitive scientists Gigerenzer, Todd et al. could be seen to be echoing one of the fundamental assumptions of relevance theory when they write: ‘Cognition is the art of focusing on the relevant and deliberately ignoring the rest’ (Gigerenzer, Todd et al., 1999, p. 21).

The disposition to search for relevance is routinely exploited in human communication, which gives rise to fairly precise expectations about how relevant an interpretation will be. Speakers know that listeners will pay attention only to stimuli that are relevant enough, and so in order to attract and hold an audience’s attention they should make their communicative stimuli appear at least relevant enough to be worth processing. This motivates the relevance theoretic comprehension heuristic below (Wilson & Sperber, 2004, p. 613):

(a) Follow a path of least effort in computing communicative effects;
(b) Stop when your expectation of relevance is satisfied.

According to the framework developed in Wilson and Wharton (2006) and Wharton (2009), pragmatic inference and relevance play central roles in the interpretation of prosody. Prosodic inputs to the comprehension process fall at various points along a continuum between natural and conventional and can be divided into four categories (see Fig. 1).
Some of the natural elements of prosody are *signs* – interpreted entirely by inference – and others are *signals*, interpreted via decoding and inferential processes. The difference between the two varieties draws on the evolutionary (and ethological) notion of adaptive function (see Hauser 1996). To use a non-verbal behaviour as an example, if Peter shivers, Mary might infer that he is cold. But the function of shivering is not to convey information: that is not the reason that shivering continues to proliferate among humans. Shivering is therefore a sign. If Peter smiles, Mary will understand that he is happy (or at least not a threat). Work in the psychology has shown that, in contrasts to shivering, the adaptive function of smiling is to convey information: it is therefore a signal (Ekman, 1999; Fridlund, 1994; Hooff, 1972). Conventionalised prosodic elements are either *linguistic signals* or *non-linguistic/cultural signals*. The paradigm archetypal linguistic signal is the word, but there are other conventional communicative signals exploited by humans according to culture-specific rules: consider, for example, the British two-fingered insult (McNeill, 2000).

These categories can be carried over to prosody unproblematically. It is not difficult to think of prosodic counterparts to the examples above. A speaker’s mental or physical state may affect the prosodic properties of her utterance, and a hearer with the appropriate experience or background knowledge will therefore be able to infer whether a speaker is sick or healthy, tired or alert. As with shivering, it is not the function of these prosodic properties to carry information about the speaker’s mental or physical state: they are natural signs, interpreted by inference rather than decoding. On the other hand, certain affective tones of voice, in a manner analogous to smiles, may well be natural signals, interpreted by innately-determined codes.

While the focus in this paper is the ‘natural’ side of multimodal communication, a word is necessary about how these innately-determined codes (characterised in Fig. 1 as natural signals) might work. In this regard, Wharton (2009) develops proposals made originally by Diane Blakemore (1987, 2002) and applied to different aspects of prosody by Vandepitte (1989), Clark and Lindsey (1990), House (1990), Escandell-Vidal (1998, 2002), Imai (1998) and

---

**Figure 19.1. Prosodic inputs**
Fretheim (2002). If, as relevance theorists take it, linguistic communication involves a combination of both decoding and inference, then coded linguistic signals might be expected to encode information of two distinct types. On the one hand there is regular *conceptual* encoding, where a word (e.g. *bird*) encodes a concept (e.g. *BIRD*) which plays a role as constituent of the logical form of sentences in which that word occurs. On the other, we might expect to find a form of *procedural* encoding, where an expression encodes information that – rather than sitting in the logical form – has the function of guiding the hearer’s inferential search for relevance. Fully linguistic expressions that have been analysed as encoding procedural information include discourse connectives and particles and mood indicators and discourse particles (cf. Blakemore, 1987, 2002; Hall, 2004; König, 1991; Wilson and Sperber, 1993).

Wharton (2003a) provides a survey of the literature and concludes that interjections – expressions such as *wow*, *yuk* and *ouch* – are best analysed as falling on the natural rather than the properly linguistic side. However, in terms of the distinctions presented above, he also argues that interjections are signals rather than signs, and that they share encode procedural rather than conceptual information. The line of argument is taken further in Wharton (2003b), where it is proposed that all natural signals should also be analysed as encoding procedures rather than conceptual information. The idea can then be extended to natural prosody, such as affective tone of voice – a natural signal – would be to facilitate the retrieval of similar propositional-attitude descriptions to those activated by interjections.

In contrast, prosodic distinctions such as the ones between *im'port* and *import*, and *ex'port* and *export*, are clearly best analysed as properly linguistic. However, parallel to the gestures referred to above, other conventionalised prosodic phenomena might be best analysed as cultural. Such phenomena might include the prosodic inputs might include the discussed by Ladd (2008). These “calling” or “hailing” contours, which are highly stylised intonation patterns, consist of two sustained notes, moving from high to low.3

Sperber and Wilson (1986/1995, 2015) have acknowledged that developing an adequate theory of utterance interpretation necessarily involves going beyond Grice’s notion of speaker’s meaning and including cases of both ‘showing’ and ‘meaningNN’. As a result, relevance theory not only allows for a much wider range of communicative acts – including, for example, the intentional showing of otherwise spontaneous natural behaviours – it also accommodates easily cases of vague communication, where what a speaker intends to convey is not a single, definite and salient interpretation, but rather a ‘diffuse impression’ (Wilson and Wharton, 2006) in the form of an array of weakly communicated assumptions (Sperber and Wilson, 2015). As Wilson and Carston (2019, p. 34) remind us:

Relevance theorists set out from the start to look for a set of pragmatic principles and mechanisms that can deal with the full range of overtly intentional communicative acts: verbal and non-verbal, showing and telling, determinate and indeterminate, literal and figurative, propositional and non-propositional.

Cases of showing and weakly communicated interpretations are often associated with non-verbal modes of communication and relevance theory fully acknowledges the role of

---

3 The archaic English *coo-eee* call is an example.
multimodal input in both context construction and utterance interpretation. Multimodal behaviours guide the utterance interpretation process by altering the salience of possible interpretations of utterances (including possible disambiguations, reference resolutions, contextual assumptions, implicatures, speech-act descriptions, etc.). What relevance theory adds to this intuitive description is the idea that the salience of interpretations can be affected not only by altering processing effort but also by manipulating expected cognitive effects. This, then, is the answer to the ‘how’ question.

Jill House (2007, p. 369) is talking about prosody when she writes that it ‘guide[s] the listener in how to proceed: how to access the relevant cognitive context within which to interpret the speaker’s contribution, how to evaluate that contribution, and how to construct the interaction itself, to enable the communication to take place’. She could, however, easily have been writing about non-verbal behaviours in general. We argue, with her, for a more natural account of prosody, which falls out from general pragmatic principles, and can also be applied to all the many modes of communication.

4. Prosody as gesture: towards an account of multimodality in interaction

As we have noted, intonation has been described as ranging along a continuum from natural to purely linguistic accounts. As a rigorous (and notorious) defender of the natural view, Bolinger (1983a) notes:

Intonation… assists grammar – in some instances may be indispensable to it – but it is not ultimately grammatical… If here and there it has entered the realm of the arbitrary, it has taken the precaution of blazing a trail back to where it came from.

Adopting this radical position, Bolinger rejects the idea of normal stress and rule-based accent placement, advocated by structuralists, on the basis that accent is only predictable if you are a mind-reader of speakers’ intentions (Bolinger, 1972). As summarised in Ladd’s words (1996, p. 167), Bolinger believed that ‘[w]hat speakers decide to highlight is […] a ‘matter of what they are trying to say on a specific occasion in a specific context’. While both accounts regard intonation as meaningful, only the natural view sees intonational meaning as essentially a reflection of speakers’ intentions and speakers’ choices as to what features of an utterance should be made more salient (Ladd, 1996, p. 167).

Bolinger’s description of a possibly pre-linguistic, almost biological, highlighting function of intonational contours used for the reading of speakers’ mental states and intentions has been disputed (Ladd, 1996). However, his description seems to suit an arguably less controversial pre-linguistic, and certainly biological, universal of human communication: pointing (Madella, 2021). Indeed, the perception and interpretation of contrastive stress in English seems to be facilitated by its co-existence and interaction with co-speech gestures, including co-pointing gestures, head movements and facial expressions. Madella (2021, forthcoming) draws on previous accounts of contrastive stress which define it as a natural highlighting device or a vocal form of pointing (Grice, 1989; Sperber and Wilson, 1986/1995; Wilson and Wharton, 2006; Wharton, 2009; Scott, 2017ab, 2021), which achieves its effects via the automatic working of the relevance-theoretic comprehension procedure. In this paper, we take this approach further and argue that prosody needs integrating into a relevance-theoretic analysis
of multimodal input. The argument for adopting a natural view of prosody indeed also coincides with a need to approach prosody as one tool in a set of broader gestural ones, and one cue in a set of contextual gestural ones.

Madella (2021, forthcoming) makes a case for acknowledging our biological predisposition to use not only prosody and gesture together, but also and more crucially, prosody as gesture. These intimate connections between vocal and gestural paralinguistic behaviours have been thoroughly investigated by Bolinger himself. In Where does intonation belong? (1983a) and Intonation and gesture (1983b), not only does Bolinger demonstrate the co-occurring of intonation and gesture, he also, and importantly, emphasises the overlapping and blending of sounds and movements in one complex of intonational and physical gestures. Reflecting this, he makes a crucial point that we ‘read intonation the same way as we read gesture’ (1983b, p.157).

This is observed in cases where ‘nods, hand gestures, and eye contact coincide very precisely with events in the spoken message’ (Ladd, 1996, p. 34; Kendon, 1972) or what Bolinger (1983c, p. 98) describes as modalities working in parallel:

If intonation is part of a gestural complex whose primitive and still surviving function is—however elaborated and refined—the signalling of emotions and their degrees of intensity, then there should be many obvious ways in which visible and audible gesture are coupled to produce similar and reinforcing effects. This kind of working in parallel is easiest to demonstrate with exclamations. An ah! of surprise, with a high fall in pitch, is paralleled by a high fall on the part of the eyebrows… A similar coupling of pitch and head movement can be seen in the normal production of a conciliatory and acquiescent utterance such as “I will” with the accent at the lowest pitch—we call this a bow when it involves the head, but the intonation bows at the same time.

The ‘bow’ phenomenon described by Bolinger can be applied to that of a nodding gesture. A nod involves a head movement, but the intonation can often be seen to nod at the same time, as if to agree, as seen in (4):

(4) He \ did

Madella (2021) follows up on this and demonstrates that we read contrastive stress or ‘vocal pointing’ the same way as we read a pointing gesture.

(5) \ He liked it.

In (5), the chin-, finger- and eye-pointing towards ‘he’ and the use of contrastive stress on ‘he’ is said to coincide. All modalities point in the same direction. This difficulty to unravel intonation contours from their paralinguistic context (Ladd, 1996) precisely shows that intonation is an integral part of the paralinguistic context.

These multi-modalities co-contribute to the speaker’s meaning. This is even more obvious when we look at a case of mismatch, or integrative interaction (Madella, 2021), as in (7):
(6) Is he trying to lose weight?
(7) He / was (forward head movement, frown, and grimace; the speaker maintains eye contact; her face shows that this is not the whole story. The use of a fall-rise (‘He \ wed’) also indicates that her answer means more than ‘he is no longer trying’.

Example (7) provides a good illustration of those fuzzy intuitions about the speaker’s meaning due to the speaker not committing to one particular interpretation (Sperber and Wilson, 2015). This second case further shows how modalities are interpreted by virtue of their interaction; thus, there are inferentially interpreted. Encoded concepts carried by individual modes of communication are to be ‘adjusted or modulated, in the course of the interpretation process’ (Sperber and Wilson, 2015, p. 145). A nod and an affirmative statement, ‘he \ was’, have to be understood in relation to the accenting of ‘was’ and the use of the past tense. We suggest that in such cases, attention to prosody can often help the hearer bridge the gap between the linguistic message and the attitude of the speaker. In the above example, the intonational input communicates that there is more to be inferred, and the speaker’s face may lead to more than one possible interpretation: he was but has given up, he was but something went wrong, she is not supposed to tell…

In English, mismatches are possible between linguistic meaning, an auxiliary verb form marked positively (e.g. ‘is’), a nodding gesture that apparently shows ‘yes’, and a frown which comes to confound the apparent positivity of the message. The paralinguistic input will be used to work out how the word is to be interpreted (Stevick, 1982). By paying attention to prosodic information, the audience understand that the speaker is disconfirming assumptions, correcting information and that behind the nodding face there is disapproval, disagreement or unhappy feelings to be found. Madella’s experimental study (2021) shows that this has particular implications for Chinese learners of L2 English, for whom ‘is’ and ‘yes’ can only mean agreement. Learning to integrate paralinguistic input in their interpretation and moving beyond encoded concepts is found to be challenging yet crucial, as it plays a central role in enhancing their oral inferential abilities and in preventing L2 fossilisation.

Scott (2021) argues that contrastive stress does not encode anything, but rather disconfirms the hearer’s expectation and acts as a cue to ostension. As such, it prompts and guides his search for extra effects for putting the hearer to more effort as a result of its unexpected pattern. One possibility is sketched below. Consider utterances of (8abc). The subtly different shades of meaning of these utterances stem solely from the contrasting position of the emboldened syllable, which is emphatically stressed and pitch prominent.

(8a) I don’t like The Beatles.
(8b) I don’t like The Beatles.
(8c) I don’t like The Beatles.

There are a range of approaches in which contrastive use of stress as typified here is analysed as part of language proper. These include the postulation of layers of information structure to accommodate it, or appeal to syntactically-determined “unmarked” and “marked” accent positions (Calhoun, 2009; Zubizarreta, 2016). Although many of these semantic analyses are intuitively appealing, it is hard to see how what are, in essence, largely descriptive proposals
can be integrated into the kind of cognitively-informed account of prosody we are aiming to develop. And there are numerous problems. The unmarked/marked distinction is controversial (Bolinger, 1972), and the notion of linguistic defaults is extremely problematic (Sperber and Wilson, 1986/95; Noveck, 2004): what is the default form in one context may not be the default form in another. A pragmatic account based on relevance theory requires no intermediate levels of description, falls out naturally from general communicative principles and can be carried over to a wide range of prosodic phenomena.

Since in English it is the final strong syllable of an intonation phrase that is accented, a hearer of (8b) or (8c) will have expectations about the prosodic contour disconfirmed. This will automatically lead them to seek an alternative interpretation, which in turn will involve extra processing effort. Given his expectations of relevance, the hearer knows that the speaker would not put him to this extra effort unless some other communicative effects were intended, the hearer is prompted to seek out an alternative interpretation. The placement of the accent naturally highlights what these alternatives might be. In (8b) the deaccenting of the syllable /biːt/ results in the preceding strong syllable being naturally highlighted. The extra processing effort the hearer has been put to will result in their forming new hypotheses about the speaker’s intended meaning: in this case, that the speaker is perhaps refuting an accusation that they don’t like The Beatles.

As another example of the strength of this kind of account, consider the proposal by Chen and Gussenhoven (2003). They argue that the interpretation of paralanguage is governed by biological codes. An example of one such code is the effort code, which links the amount of energy expended in the production of speech to a range of interpretive effects. An increase in effort may, for example, lead to increased articulatory precision, creating an impression of ‘helpfulness’, or ‘obligingness’; or it may result in a wider pitch range, creating an impression of ‘forcefulness’ or ‘certainty’ or conveying affective meanings such as ‘agitation’ or ‘surprise’. But is clear articulation a natural signal, interpreted by an innately determined biological code? We suggest that it is much better treated as a natural sign of the speaker’s desire to be understood and interpreted via inference rather than decoding. This particular element of prosody belongs under the left-most node (see Figure 19.1). A great deal can be communicated by using clear articulation – that you intend to convey helpfulness, or that you are being obliging, or that you want to convey any one of a wide range of other impressions – but nothing is encoded by it at all.

The account developed by Madella (2021, forthcoming) lends further support to Scott’s account (2021) by suggesting that the co-occurring modalities reinforce the presumption of extra effects which in turn are retrieved by attending to contrastive stress and how it relates to its co-occurring counterparts. It is precisely because it is never interpreted in isolation but, rather, in the context of its co-occurrence with other modalities, that contrastive stress can be said to be a non-encoded phenomenon. It is by virtue of its interaction with co-pointing modalities, either pointing in the same direction or to the fact that there is more to infer, that it is inferentially interpreted. It is, as we argue here, one prosodic tool in a larger gestural complex, or what Madella calls prosodic pointing (Madella and Romero-Trillo, 2019; Madella, 2021; Madella, forthcoming).

This forms the biological argument for a multimodal view of communication. Gesture researchers describe those concomitant paralinguistic elements as ‘parts of a single psychological structure’ (McNeill, 1985, p. 351). Gesture researchers (Kendon, 1980, 2004, 2014ab; McNeill, 1992; Clark, 1996) agree that vocal and visual modalities together form the act of communication, namely the utterance. It is in line with the immediacy assumption (Hagoort and van Berkum, 2007, p. 808), according to which we, communicators, take ‘all
communicative acts, including eye gaze, gestures, smiles and pointing, into consideration’ and process them immediately. The immediacy assumption is based on the observation that the mind follows the eye (Hagoort and van Berkum, 2007). This argument also resonates with Claudel’s description of total and immediate receptivity in his essay *L’Oeil écoute* (1946). This is also in line with the integration of paralanguage into a general view of communication. For example, conversation analysts Streeck et al. (2011, p. 7) describe multimodal interaction along these lines: ‘Talk and embodied behaviours occur as interdependent phenomena, not separable modes of communication.’

It follows that it is by virtue of their interaction that those interlinked verbal and non-verbal behaviours constrain and guide the hearer in reaching an interpretation. It is also by virtue of their interaction that encoded concepts carried by individual modes of communication are to be ‘adjusted or modulated, in the course of the interpretation process’ and for the purpose of that one particular interpretation (Sperber and Wilson, 2015, p. 145). An inclusive definition of paralinguistic phenomena is also and precisely what allows us to argue for an inferential model of (verbal and non-verbal) communication from the perspective of relevance theory.

5. Conclusion

Wharton (2009, p. 1) begins as follows:

Sentences are rarely uttered in a behavioural vacuum. We colour and flavour our speech with a variety of natural vocal, facial and bodily gestures, which indicate our internal state by conveying attitudes to the propositions we express or information about our emotions or feelings. Though we may be aware of them, such behaviours are often beyond our conscious control: they are involuntary or spontaneous. Almost always, however, understanding an utterance depends to some degree on their interpretation. Often, they show us more about a person’s mental/physical state than the words they accompany; sometimes, they replace words rather than merely accompany them.

The statement is not less true now than it was then. People working in pragmatics cannot, as many theoretical linguists do, abstract away from these behaviours. In this chapter we have introduced the major role non-verbal modes play in communication and, in so doing, attempted to promote a ‘multimodal’ view of human communication and show how certain behaviours guide our interlocutors to our intended meaning by constraining the construction of context. We have answered three questions:

1. What is multimodality?
2. How do non-verbal behaviours communicate?
3. How can we move towards a principled pragmatic account of multimodality in interaction?

We argue that a pragmatic account based on relevance theory can be carried over to a wide range of multimodal phenomena, requires none of the intermediate levels of description that
necessarily form part of less pragmatic accounts and falls out naturally from general communicative principles. There are many different accounts of multimodality (Kress and van Leeuwen, 2001; Forceville, 2021), but the blade of Occam’s Razor, we argue, favours the one presented here.

References


