

THE BOUNDARIES OF LANGUAGE: DEALING WITH PARALINGUISTIC FEATURES

MACIEJ KARPIŃSKI

ABSTRACT: Maciej Karpinski. *The Boundaries of Language: Dealing with Paralinguistic Features*. *Lingua Posnaniensis*, vol. LIV (2)/2012. The Poznań Society for the Advancement of the Arts and Sciences. PL ISSN 0079-4740, ISBN 978-83-7654-252-2, pp. 37–54.

The paralinguistic component of communication attracted a great deal of attention from contemporary linguists in the 1960s. The seminal works written then by Trager, Crystal and others had a powerful influence on the concept of paralinguage that lasted for many years. But, with the focus shifting towards the socio-psychological context of communication in the 1970s, the development of spoken corpora and databases and the significant progress in speech technology in the 1980s and 1990s, the need has arisen for a more comprehensive, coherent and formalised – but also flexible – approach to paralinguistic features. This study advances some preliminary proposals for a revised treatment of paralinguage that would meet some of these requirements and provide a conceptual basis for a new system of annotation for paralinguistic features. A range of views on paralinguistic features, which come mostly from the fields of speech prosody and gesture analysis, are briefly discussed. A number of assumptions and postulates are formulated to allow for a more consistent approach to paralinguistic features. The study suggests that there should be more reliance on continua than on binary categorisations of features, that multi-functionality and multimodality should be fully acknowledged and that clear distinctions should be made among the levels of description, and between the properties of speakers and the speech signal itself.

Maciej Karpinski, Institute of Linguistics, Adam Mickiewicz University, al. Niepodległości 4, PL – 61-874 Poznań, maciej.karpinski@amu.edu.pl.

1. MOTIVATION FOR THE PRESENT STUDY: DEFINING THE PROBLEM

In corpus-based studies of spoken language, linguists face a number of practical problems caused by the apparently distant and abstract issues of the definition of language, the understanding of its boundaries and the concept of paralinguistics. This section provides a brief introduction to the notion of paralinguage and paralinguistic features and states the general aim of the study.

Throughout the last century and continuing into the present one, our understanding and representation of language as a system has undergone a number of changes, some of which have been dramatic. When considering the complexity of language, linguists have often made attempts to express its mechanisms in the simplest possible terms, such as Chomsky's Minimalist Programme (CHOMSKY 1993), as well as attempts to simplify or reduce

the set of features that define it (HAUSER et al. 2002). At the same time, models of language mechanisms have become more flexible and shifted from algorithmic, one-way approaches towards connectionist representations based on networks, probability and fuzzy logic. These may be less intuitive but probably better reflect the actual language-related and highly contextualised mental processes (for a review of approaches to language modelling, see DIJKSTRA & DE SMEDT 1996; GARROD & PICKERING 2000; BOLSHAKOV & GELBUKH 2004; CLARK et al. 2010). However, while the descriptive power and flexibility have increased, most of the new approaches still tend to suggest a boundary between what belongs and what does not belong to the “language system proper”. The definition of language remains arbitrary. Linguists realise that they extract their objects of interest from an extremely complex tissue of multi-level social, cultural and psychological factors but are sometimes not able to predict the full consequences of such steps.

With the growing attention paid to language in use and interaction, and to the actual realisations of utterances, the question of well-formedness of their components can no longer be ignored. Presumably some of them can be regarded as realisations of abstract linguistic units that “reside in minds” in the way that phones can be regarded as realisations of abstract phonemes. However, some of the salient properties or components of spoken utterances do not appear to correspond to any recognised units of the language system as this is usually understood. Yet they still contribute to the meaning of utterances. They may provide information on the speaker, his/her social background, his/her attitude to the topic or his/her conversational partner as well as on the entire communicational situation. Within a given culture and language, many of these phenomena are commonly understood as expressions of attitudinal and emotional states or as indicators of the speaker’s condition, such as joy, sadness, tiredness or hesitation. Some of them also seem to have a cross-linguistic or cross-cultural meaning (e.g., ABELIN & ALLWOOD 2000; BURKHARDT et al. 2006; THOMPSON & BALKWILL 2006). This applies to a large extent to both realisation-related features, such as “melody” or “voice quality”, and to the components of utterances themselves. The quasi-lexical units that occur in speech, such as “uhm”, “um” and “m”, may be difficult to find in dictionaries but are certainly frequent in colloquial speech. They tend to bear easily recognisable meanings – while remaining sensitive to prosodic override – and can hardly be classified as grammatically ill-formed, even though their realisations often barely satisfy the phonotactic rules (KARPIŃSKI 2007). Moreover, when a spoken utterance is produced it occurs alongside other kinds of behaviour, such as gestures and changes in facial expression or posture, which may significantly influence the meaning.

As all these “peripheral” phenomena have often been understood as not belonging to language proper but existing somewhere close to it because of their influence on the meaning and their overall role in communication, they are often collectively termed “paralanguage” (HILL 1952, 1958). If the aim is to achieve a holistic and integrative approach to communication, these phenomena cannot be ignored. To this end, a number of researchers have attempted to devise ways of treating them more systematically (e.g., spoken language grammars (CAMPBELL 2002), grammars of gesture (FRICKE et al., in prep.), or multimodal grammars (FRICKE 2012; ALAHVERDZHEVA et al. 2012)).

Even some relatively early works on the sonic substance of language postulated a surprising flexibility and openness in the understanding of what should be included in lan-

guage or what at least would be impossible to neglect in linguistic analyses (e.g., PIKE 1943; TRAGER 1958). Seminal studies by ABERCROMBIE (1967), CRYSTAL & QUIRK (1964), CRYSTAL (1966), SMITH (1952, 1953) and TRAGER (1958, 1960, 1961) still remain very influential and provide the conceptual framework for more recent views of paralinguage. Nevertheless, one cannot overlook that much has also changed. The growing interest in spoken language, phonetic variation, as well as in the social and psychological aspects of communication, gave a new impulse to paralinguistic studies in the 1970s. The study of the paralinguistic component gained further momentum in the following decade with the rapid development of computer-based speech synthesis, speech recognition and speaker identification. While these studies went into ever deeper phonetic detail and used increasingly complex statistical measures, they may have neglected some conceptual aspects of the problem. As a consequence, the ways in which the paralinguistic component of communication is currently understood varies from one author to another to a surprising degree. This may indicate that the conceptual framework for the study of paralinguistics requires revision to restore coherence, needs to adopt more current views of language and should meet the new requirements for the description of language data.

The present text has been prompted by the very practical needs of spoken language corpora annotation and analysis. Its intention is not to propose a theoretical critique or reformulation, but to provide some cues and suggestions regarding how the paralinguistic component of communication should be treated in order to preserve the meaning it can carry and to appreciate its communicational vitality.

2. PARALANGUAGE AND PARALINGUISTIC FEATURES: DEFINITIONS AND PROPERTIES

As shown in Section 1, linguists have been conscious of the problematic boundaries of language and the importance of the phenomena collectively referred to as “paralanguage” for quite a long time. On the other hand, the literature is striking in the diverse ways it understands, defines and interprets paralanguage and paralinguistic features (PFs), that is, the features of utterances that have no relevance to “language proper” as it is currently understood.

In a number of studies, paralinguistic phenomena have been viewed as bearing significant resemblance to what can be observed in animal communication. David Crystal claimed that some nonverbal vocal behaviour “is held to be a significant point of overlap between human and animal communicative systems” (CRYSTAL 1976). He refers to STURTEVANT’S (1947) claim that the exclamatory parts of language closely resemble the sounds produced by many animals because of their extreme changes in loudness and pitch. Further, he refers to the work of ABE (1967) and THORPE (1972) where the terms “paralanguage” and “paralinguistic” are used to refer to animal communication. He also points to the fact that the notion of “paralanguage” has been “more of a hindrance than a help to progress in our understanding of nonverbal and vocal behaviour” (CRYSTAL 1976: 13). It is difficult to disagree, as in fact the boundaries between the linguistic, paralinguistic and non-linguistic turn out to be only superficially cognitively comfortable and their fuzziness causes even more confusion.

PFs can also be regarded as “indexicals”, that is, discourse components that indirectly refer to the position and background of the speaker and can be interpreted as such by

talk participants (GUMPERZ 1992). The notion of indexicality, which was introduced to linguistic analysis by SILVERSTEIN (1976), refers to the way in which the meaning of every utterance is based on how it was produced and what it says about the speaker. Although Silverstein's work is directly inspired by Pierce, a similar concept can be traced back to Jakobson's studies of sound patterns. He also mentions "shifters" or signs that may change the meaning depending on the identity of the speaker and the context (JAKOBSON & WAUGH 1979). GUMPERZ (1992) defines "contextualisation cues" as features of speech that relate what is said in a given time and place to past experience. Such cues are grouped into three categories: phonological (intonation, stress, pitch), paralinguistic (gestures, facial expressions) and linguistic (choice of code, choice of lexical forms, choice of formulaic expressions). Here, the linguistic status of PFs is clearly denied and they are limited to body motions.

CRYSTAL (1969) introduced the categories of "prosodic" and "paralinguistic" features and claimed that "while prosodic features are a necessary component of all speech, PFs may be absent, and allow for more idiosyncrasy in their realisation". Again, depending on the adopted definition, it may turn out that it is actually impossible to produce an utterance that is completely without PFs. All utterances, even computer-generated ones, have a particular rhythm (even if monotonous), intonation (even if flat) or voice quality (even if not very striking) that is "non-linguistic" in the traditional sense. CRYSTAL & QUIRK (1964) distinguished two categories of PFs: (a) voice qualities that result from different phonation modes (normal voice, falsetto, whisper, creak, huskiness and breathiness) and (b) voice qualifications that are "non-linguistic vocal effects running through or interrupting speech". These include laughing, giggling, tremulousness, sobbing, and crying. This distinction seems to hold quite well as it is based more on the technical properties of speech itself. One may nevertheless expect crying or laughing, for example, to correlate with quite particular, dynamic changes in phonation.

PFs can be also understood as vocal features which are independent of the traditionally defined segmentals and supra-segmentals and whose presence is not obligatory (BROWN 1990). This approach also has some weaknesses. One may argue, for instance, about the degree of independence. First of all, even if one assumes that for each utterance there exists a kind of linguistic base on which PFs are somehow "layered", this base may determine what can actually be added and built on to it. In an extremely short utterance, tempo variation will probably not be an effective means of expressing emotionality because there may not be enough space to make this variation perceptually salient. Secondly, the "linguistic" component may be strongly influenced by the factors that form the "paralinguistic" component. It is not difficult to observe that emotions may not only cause a change in voice characteristics but may also influence the choice of vocabulary and syntactic structures. On the other hand, as was mentioned above, it is difficult to imagine an utterance with no PFs.

There have been a variety of interpretations and applications of the term "non-linguistic", some of which have been close to those of the term "paralinguistic". For example, LAMEL & GAUVIN (1993) use it to refer to "speech features that give little information about the linguistic content of the message, such as the gender or identity of the speaker, the accent or even the language spoken". While interesting for its form and innovativeness, there are two major reasons why this definition is not fully satisfactory. First of all, it is not very pre-

cise when it comes to the actual allowable amount of linguistic information (“little information” does not mean “no information”). Secondly, it assumes knowledge of “the linguistic content of the message” which, as one may infer from the brief discussion above, is neither clear nor trivial.

In his discussion of computer-mediated communication, CAREY (1980) accepts both the vocal features outlined by TRAGER (1964) and the prosodic system of CRYSTAL (1969) as components of paralinguage. Concerning Trager and Crystal he points out that “both are concerned with the investigation of linguistic phenomena which generally fall outside the boundaries of phonology, morphology and lexical analysis. These phenomena are the voice qualities and tones which communicate expressive feelings, indicate the age, health and sex of a speaker, modify the meanings of words and help to regulate interaction between speakers” (CAREY 1980: 67). In this way “paralanguage”, as understood by Carey, remains a “linguistic phenomenon”!

It has sometimes been assumed that any non-verbal component of an utterance is paralinguistic. For example, DAVIES & WIDDOWSON (1974) declared that the paralinguistic component corresponds to the non-verbal one in face-to-face communication, which means mostly gestures and tone of voice. However, as early as 1985, McNEILL was questioning the tradition of referring to gestures as “non-verbal”. In the recent reading of ROSSINI (2012), meanwhile, some varieties of gestures may have a very precise lexical meaning or at least may be considered as having “lexical access”.

Finally, there is near-universal agreement that linguistic features are discrete. This view is expressed by, for example, VON HEUSINGER (1999: 17) (“something is either a /p/ or a /d/, an utterance has either a question intonation or an assertion intonation, the focus sensitive particle associates with this or that expression, etc.”) and, with special reference to prosody, by GUSSENHOVEN (2002; see 2.1 below). One may still expect such a distinction also to be useful with reference to other kinds of behaviour, such as gestures. On the other hand, one may suspect that not all of the categories distinguished in linguistics are fully justified and that they are based more on empirical assumptions than on the preconception of a language system as a categorical one.

PFs are discussed below in more detail in the context of studies of prosody, voice quality and gesture. Studies have frequently addressed paralinguage in these three areas as they present the greatest challenges and have made the greatest claim on the attention of researchers. Some of the achievements of phonetics and gesture studies can easily be employed to obtain a better view of paralinguage or can, at the very least, provide inspiration when attempting to find solutions to more general problems.

2.1. PROSODY AND PARALINGUISTICS

As far as the phonetic level of analysis is concerned, prosody is typically described as involving length, accent, intonation and tone (CRUTTENDEN 1986; FOX 2000). This can also be reduced to three basic parameters of pitch frequency, duration and intensity. These are quasi-independent, which means that they can be manipulated almost independently, but normally have a strong influence on each other. The issue of the boundary between the linguistic and the non-linguistic remains especially prominent in studies of prosody, which may be a rich source of information on various language subsystems. CHEN (2005: 9), for ex-

ample, proposes the following inventory of “intonational meanings”: grammatical meaning (speech acts), pragmatic meaning (pragmatic uses of speech acts), attitudes of the speaker and discoursal meanings (“the relation between the variable in focus and an implicit set of variables in the context.”).

It would be risky to claim that intonational contours themselves have (referential) meanings (CUTLER 1977). Still, prosody is employed to convey a wide range of “meanings” using various dimensions of the speech signal. While speech rhythm contributes immensely to intelligibility (e.g., HUGGINS 1979), intonation would appear to be the most predisposed and efficient bearer of complex meaning. To a large extent (but not solely!) pitch changes contribute to lexical meaning (e.g., stress, accent and tone), convey phrase-level information (e.g., sentence category, phrase boundary), provide cues to discourse functions (e.g., segmentation, prominence, focusing, topicalisation), add affective quality (e.g., LIEBERMAN & MICHAELS 1962; HIRSCHBERG & PIERREHUMBERT 1986; CORNISH 2005; MOZZICONACCI 1995; HIRSCHBERG 2002; BÄNZIGER & SCHERER 2005) and offer many other types of (Piercean) indexical information, such as nationality, age or background (CRYSTAL 1987; see also, e.g., ABERCROMBIE 1967; consult FOX 2000 for an insightful review of the functions of prosody).

The question of how any changes in the signal can be unequivocally attributed to the linguistic or the paralinguistic factors is a serious one when so many tiers of meaning are encoded in those few prosodic dimensions. One possible approach to this problem, advocated by, for example, GUSSENHOVEN (2002, 2004), is to clearly distinguish between intonational phonology and morphology, and the phonetic implementations of intonation. The former are language-specific while the latter follow more general tendencies. GUSSENHOVEN (2002: 50–57) argues that the linguistic component of intonation is characterised by arbitrariness, discreteness and duality. He attributes the paralinguistic use of pitch to three biologically determined factors which are reflected in three Biological Codes: the Frequency Code, introduced earlier by OHALA (1983), the Effort Code and the Production Code. These codes are divided into attributes of the message and attributes of the speaker. While this idea provides a relatively coherent view of intonational phenomena, it turns out to be difficult to implement in the practise of intonational analysis, as the hypothetical components intervene and mix in each utterance. In Polish, for example, a final rising tone may indicate the linguistic category of “question” or the emotional category of “astonishment”, or can be used consciously as a stylistic means to attract attention, to raise tension or to signal planned continuation (KARPIŃSKI 2006; KLESSA & KARPIŃSKI 2012). Segment lengthening may be “phonological” or occur just as an expression of emotionality or stylistic manipulation. High pitch may be an intrinsic parameter of a child’s voice resulting from the size of the glottis. In an adult male, though, it may mark a highly emotional expression or a joking register. But it can be also be used intentionally by adults to express such things as subordination (OHALA 1983; GUSSENHOVEN 2004).

A further step is taken by HIRSCHBERG (2002), who extends the Gricean set by proposing additional conversational maxims based on the Biological Codes proposed by GUSSENHOVEN (2004). The IPO (Institution for Perception Research) school argues that intonational features have no intrinsic meaning (T HART et al. 1990) while CHEN (2005), referring further to HAAN (2001: 33–35), points out that “it has proved possible to uncover intonational meaning by excluding lexical and contextual intonation”. CHEN et al. (2004: 312) claim that “it is gen-

erally accepted that paralinguistic uses of intonation are remarkably similar across languages. That is, while languages can and do differ widely in how discrete pitch categories are used to contrast lexical or phrasal meanings, all languages seem to use gradient variations in pitch properties to signal different degrees of a certain set of meanings". Still, as shown by ABELIN & ALLWOOD (2000), for example, emotional prosody can be, at least to a certain extent, language-specific and culture-specific. CHEN et al. (2004) point out that there are language-specific uses of the Effort Code as defined by GUSSENHOVEN (2004). WICHMANN (2002) mentions that the attitudinal cues to intonation are also "known as language-specific". BURKHARDT et al. (2006) conclude their study conducted in France, Germany, Greece and Turkey with a tentative statement that emotional prosody is largely universal, but that it still has some language-specific and culture-specific components. According to MOZZICONACCI & HERMES (1997), it is not only pitch level and the range of pitch changes that can be regarded as cues to emotionality because there are also certain intonational patterns typical of emotional utterances. While this is insufficient to support a "grammar of emotionality", it certainly provokes more general thoughts on the patterns of communicative behaviour and on the criteria for "being linguistic".

In most situations, paralinguistic prosody can be consciously managed by the speaker to express emotional or other kinds of meaning in a controlled way. However, on other occasions the same aspects may prove difficult to control due to the actual emotional state itself. Obviously, speakers vary in their ability to control prosody and, primarily, to hear it consciously.

This brief review clearly shows the potential of prosody as a means of conveying both linguistic and paralinguistic meanings (if we are still prone to adopt this division). Both large-scale prosodic phenomena, such as rapid pitch change, and more subtle ones, such as precise pitch-peak alignment, may carry meaningful information. Each acoustic dimension can be used to convey both linguistic and paralinguistic features. Furthermore, it is extremely important to state here that the gradual vs. discrete distinction is not the property of the acoustic features themselves but rather of the mechanisms of perception that are just as prone to impose categorisation – even if there are no strictly acoustic cues to categories.

2.2. VOICE QUALITY AND PARALINGUISTICS

Voice quality is considered to be even more individual, idiosyncratic and further from "language proper". Nevertheless, it may also have phonological relevance and language-specific usage. According to ABERCROMBIE'S definition (1967: 91), "voice quality refers to those characteristics which are present more or less all the time a person is talking: it is a quasi-permanent quality running through all the sound that issues from the mouth". LAVER (1980: 1) and TRASK (1996: 381) defined voice quality as the "auditory colouring" of an individual's voice which is caused by a variety of laryngeal and supra-laryngeal factors spread across spoken utterances. In this sense, it is close to a wider notion of *timbre* (as applied to sound in general, not only to speech).

It is evident that voice quality, as defined above, is multidimensional and difficult to operationalize. KREIMAN et al. (2004) point to the fact that "the overall quality of a sound is formally defined as that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar".

They also argue that many popular definitions of voice quality are negative and focused on what voice quality excludes without precisely listing what it includes. It is even more elusive when defined in terms of perception, as this is influenced by many individual, social and cultural factors. Some qualities of voice tend to be described not in terms of auditory impressions but rather by referring to certain features of the speaker or his/her behaviour (a *kind* voice, an *angry* voice). Of course, even more direct descriptions of voice quality often remain metaphoric.

Voice quality is sometimes referred to as the “fourth prosodic dimension” (CAMPBELL & MOKHTARI 2003). Still, the measurements of acoustic correlates of voice quality can be based on relatively short samples of speech, not as long those often necessary for prosodic analysis. One also cannot expect voice quality to remain perfectly stable throughout an utterance. Instead it may evolve, especially with emotional changes or due to the sudden impact of external factors on the speaker.

Some components of voice quality, such as harshness, are clearly language (culture?) dependent, while in the case of others, such as breathiness, the difference does not seem significant (GHIO et al. 2011; see also KEATING & ESPOSITO 2006). Such findings are, on the one hand, unsurprising, because some phonetic features may contribute to phonology in one language and not necessarily in others. Cerebral consonants may function as a phonological category in a given language (e.g., Hindi), while, in other languages, using cerebral articulation may add some attitudinal value to an utterance, but will not change the phonological status of a consonant.

Emotional states may heavily influence voice quality, and many of these influences may be universal (e.g., SCHERER 2003; SCHERER et al. 2001; GOBL & CHASAIDE 2003; ISHI et al. 2006; LUGGER & YANG 2007, 2008). Nevertheless, one may still expect some language-specific effects and one may suspect that the results presented by ABELIN & ALLWOOD (2000) for cross-linguistic perception of prosody can probably also be obtained for cross-linguistic perception of voice quality.

It has been shown that the difference between male and female voices resolves itself not only in pitch but is strongly reflected in many other aspects of voice quality. A detailed account of the gender-related voice quality variation was given by BIEMANS (2000). Voice quality may cue some personality traits but there have also been some attempts to use it to diagnose psychiatric disorders (CHEVRIE-MULLER et al. 1978). LAVER (1994: 22) mentions an example of cultural differences in the falsetto voice. If it is used in English to mimic a male conversation participant’s utterance, it means an accusation of effeminacy, while in Tzeltal (when used in greetings) it serves as a marker of deference.

The most comprehensive phonetic treatment of voice quality has probably been offered by Laver and his collaborators. LAVER & TRUDGILL (1979) characterise paralinguistic voice features as “informative” but not “communicative”. They distinguish between two types of long-term voice features: anatomically induced voice characteristics and vocal settings (phonatory, articulatory, tension and prosodic; LAVER 1994). Voice characteristics are determined by anatomical factors, such as the size of the larynx and body mass, which cannot be changed by the speaker without a serious amount of effort. Vocal settings can be changed at will although they are not consciously manipulated all the time. One may suspect that a given shape and condition of the articulatory tract somehow also influences the individual

use of vocal settings. LAVER (1980) also provides a comprehensive study and categorisation of voice-quality components, which is referred to as Laver's Voice Quality Schema (KELLER 2004). What is especially important here in the light of the present study is the distinction between anatomically-determined and volitional parameters. The former can perhaps be likened to a speaker's fingerprint, but the latter may prove equally, or even more, important in determining a speaker's communicational and psychological characteristics. Another idea of practical importance that occurs here is the distinction between the long-term and short-term features, which seems relevant at the acoustic and kinetic levels as well as at the more abstract level of analysis.

2.3. GESTURES AS PARALANGUAGE

As mentioned above, the term "paralanguage" may cover not only some aspects of speech but also of body motion – especially, gestures and facial expressions. BAVELAS (1994) stresses that gestures are part of speech, while ROSSINI (2011) goes even further in claiming that they should be regarded as part of language. CASSEL et al. (1994), meanwhile, have argued that gestures are not simply "translations of speech" but nor are they irrelevant to it. Moreover, there is a growing body of evidence that gesture and speech come from the same internal, intentional source (ARBIB 2005; MCNEILL et al. 2005; see KARPIŃSKI et al. 2009 for a brief discussion). Gestures, therefore, may lie closer to each other than previously thought and, as a consequence, may not be far from "language proper". But gestures themselves differ to a significant extent in terms of their form, meaning and usage. In the history of gesture studies, a number of gesture categorisations have been proposed and applied (KENDON 2005: 84–112). EKMAN & FRIESEN (1969), for example, introduced the categories of (a) emblems, (b) illustrators, (c) affect displays, (d) regulators and (e) adaptors. Currently, however, researchers tend to admit that it is often difficult to define clear boundaries between neighbouring categories as they seem to form a continuum on "the axis of linguisticity". This is often referred to as Kendon's continuum (MCNEILL 1992):

Gesticulation → Speech-Linked → Pantomime → Emblems → Sign language

When elaborating their view of gesture typology, inspired by Peircean semiotics, MCNEILL & LEVY (1982) referred not to categories but rather to dimensions, which include iconicity, metaphoricity, deixis and beat. All lie relatively close to the left end of Kendon's continuum (gesticulation or speech-linked gestures). MCNEILL (2006) claimed that gesticulations combine both universal and language-specific features. In this context, their proximity to prosody and voice quality seems to be even more evident – and more analogies can be found.

While gesticulation is driven mostly by current meanings, not by conventions, it is still prone to adapt to the form of the co-occurring spoken utterance and to reflect its informational or discursive structure. Gestures can successfully replace words or even phrases in speech. At the same time, and probably first of all, gestures mirror speakers' emotional states and enhance or change the meaning of spoken utterances.

It has been shown that gestures are often produced more for the speaker than for the listener (the external addressee), as they help in the process of conceptualisation and, as a consequence, in speech production (JARMOŁOWICZ-NOWIKOW & KARPIŃSKI 2011). When all

the similarities between gestures and voice-incorporated paralinguistic features are taken into account, one may hypothesise that voice quality and prosody may function in a similar way, not only in enriching the informational value of utterances but also in facilitating the process of expressing intended meanings.

The tools of linguistics turn out to be surprisingly appropriate for the study of gestures. GIBBON (2011) stresses the similarities between gesticulation and prosody and shows how linguistic methodology can be applied in gesture analysis. The patterns of gesticulation and prosody are, as a rule, time-aligned and perform similar functions, such as making some portions of utterances more prominent (YASINNIK et al. 2004; CUMMINS & PORT 1996). Therefore, the stream of gesticulation may be easily classified as a bearer of paralinguistic meaning. The striking similarities and alignments between prosody and gesticulation concern both form and function. Gesticulation is probably a means of colouring the meaning of utterances with attitudinal states and personal traits. But certain categories of gestures can also replace words or utterances and have an independent meaning. As in the case of prosody and voice quality, we may argue whether gestures (when? which?) should be incorporated into the language system proper or whether we should continue to regard them as peripheral to it.

3. TOWARDS A COHERENT APPROACH TO PARALINGUISTIC FEATURES

While over forty years have passed since ABERCROMBIE (1967: 5) admitted that extra-linguistic properties might fulfil certain functions that sometimes prove more important than linguistic communication itself, the status of paralinguistic and paralinguistic features remains rather ambiguous in contemporary linguistics and communication studies. It is still difficult to draw a detailed, comprehensive and coherent image of paralinguistic and perhaps even to clearly justify the existence of the term itself. If so many entities that are very distant from “language proper” can be included in this category, why should they be referred to as paralinguistic? How “close to language” should a phenomenon be to remain paralinguistic? How strongly should it be related to the “language proper”? Or is it perhaps the case that this division is retained only for the sake of “cognitive comfort” and in order to support the *status quo* of linguists and linguistics? There may be an underlying assumption that language is central to communication and that the rest of the phenomena under discussion merely surround and support it. Yet it is difficult to deny that the “linguistic core” of a message does not always bear the most important part of the message.

These and other fundamental questions are definitely beyond the scope of the present study as it is intended merely to suggest some directions to be considered on the way to a more coherent model of communication and to alleviate some practical problems faced in the realm of corpus-based research. One may have to assume that the paralinguistic component of communication must remain under-specified – as is in its nature – and that the boundary between “the linguistic” and “the non-linguistic” must remain arbitrary. On the other hand, one may wonder if such a division should exist at all. Some phenomena in interpersonal communication are more easily conceptualised as “systematic” or “structuralised”, while others are more difficult to categorise in this way. It may be remembered, though, that the traditional descriptive categories of linguistics had emerged well before the full acknowledgement of the social and psychological factors at stake and, also, well before more

comprehensive attempts at viewing communication as a dynamic, multimodal process, such as those recently proposed by Cornelia Müller, Alan Cienki and their colleagues, were made.

It is proposed here that the paralinguistic component of communication be understood as all the phenomena and features of a speaker's behaviour that go beyond the (current) limits of systematic linguistic description but still influence the way his/her communicational contribution is understood by his/her conversational partner(s). This includes conscious, intentional performance but also those features of articulation, facial expression or gesticulation that are not fully controlled by the speaker. These can be perceived not only as a display of attempts to communicate but also as cues to the mental and physical state of the speaker. It is also possible to take a more speaker-oriented perspective but, as the criterion of intentionality is abandoned anyway and, as in the analysis of the process of communication one has not much insight into the mind of the speaker, the efficiency of such an approach is disputable – even if it certainly has some value and may be elaborated in the future.

The definition proposed above may be criticised as having an extremely wide reference. Nevertheless, it achieves the aim of gathering a number of heterogeneous phenomena under one umbrella, while refraining from detailed description or enumeration of the paralinguistic features or components. In this way, the arbitrary exclusion of any potential phenomena is avoided. Its scope is not limited to what is most commonly taken into account in interpersonal communication studies and it is inspired by various attempts aimed at a holistic description of communicative events, such as those of HYMES (1974).

With this working definition in hand, and based on the discussion that has encompassed the fields of speech prosody, voice quality and gesture studies, the following five recommendations are made that can help us to grasp the importance and particular properties of the paralinguistic component of communication – both in the theory and practise of linguistic data annotation.

3.1. CONTINUA RATHER THAN CATEGORIES

As mentioned in 2.1, GUSSENHOVEN (2004: 51–53) describes paralinguistic intonation as non-arbitrary and lacking in discreteness and duality. One may definitely argue with his proposals and with the traditional assumptions about the properties of the linguistic categories themselves that he employs in his study. Nevertheless, the criteria introduced by Gussenhoven for intonation may well be applied to other channels and modalities used in communication. When viewed from this particular viewpoint, paralanguage is extremely flexible but perhaps lacking informational and expressive precision. Gestures based on similar motion patterns but varying in their physical range (magnitude) or velocity may carry, for example, a whole range of (non-discrete) emotional meanings, such as a varying amount of joy or anger. But the general pattern of the motion trajectory may remain unchanged and convey the same “core” meaning, as in the case of iconic gestures.

Kendon's continuum (see: 2.3) may also inspire a new approach to the potential categorisation of paralinguistic phenomena. He suggests that gestures do not cluster in mutually exclusive categories on the axis (dimension) of linguisticity, but are rather spread across it in fuzzy groups. Accordingly, one may assume that various categories of paralinguistic features may be closer to or further from language and thus abandon the idea of rigid boundaries and the criteria of linguisticity.

Lack of discreteness would seem to be typical of most paralinguistic features. But paralinguistic annotation systems rarely offer the means to reflect this fact and compel annotators to operate on coarse-grain or just binary categories, such as harsh vs. soft. It is also important to notice that it is not only the continuity of the features *per se* that is neglected, but also that the extra-linguistic categories used to label corresponding states of the speaker are imposed as discrete. This is the case with affective states, such as joyful vs. sad (ROACH 2000), for example. One possible practical solution is to implement virtual sliders or joy-sticks for data input and to refrain from imposing any points on the scale, such as from a stable to a trembling voice, with all intermediate states possible (KLESSA & KARPIŃSKI 2012).

3.2. MULTIFUNCTIONALITY

Contemporary approaches to the semantics of natural conversation are dominated by the assumption that utterances (or conversational contributions) are multifunctional (BUNT et al. 2010; BUNT 2011). PFs, both as individual entities and as certain sets clustered together by the mechanisms of perception, also seem to function on multiple levels. A single feature (or a set of features) can bear multiple meanings or realise a number of functions. For example, a filler such as “um” may signal a pause required for thinking and, simultaneously, express the mood of the speaker, or even his or her social status or origin (if there are local variants of its pronunciation). Care should therefore be taken when annotating fillers on the functional level to distinguish all of their relevant functions. Obviously, not all of them are equally important in a given context, which can also be reflected in the annotation as has been suggested, such as at the level of dialogue acts (WŁODARCZAK 2012).

3.3. MULTIMODALITY

PFs in speech, gestures and facial expressions are nothing but bodily movements or their acoustic consequences (GIBBON 2011). Humans perceive them through various modalities, such as hearing, vision and touch. As mentioned above, it has been widely argued that the final, multimodal form of the message, including both the articulatory and other body movements (e.g., gestures), normally originates from a single intentional source. Various modalities share at least some neural circuitry activated in the process of producing an utterance. Nevertheless, each modality has its particular characteristics, which are partially related to its physiological limitations. Moreover, it is not infrequent that one modality is better controlled and provides information that is quite contradictory to that shown by the other modalities (e.g., the verbal content of an utterance indicates joy while the face of the speaker expresses sadness). This may be a good reason to describe the respective modalities on separate annotation tiers. Furthermore, there can also be quasi-independent channels within each modality as is the case with verbal and intonational units in speech, which pass through the auditory channel at the same time, but which carry potentially independent meanings (KARPIŃSKI 2006). However, the resulting impression, based on all the streams of information, should also be tagged, even though it may be more individual and subjective than the perception of separate modalities. In this way, it will be possible to re-construct the impact and meaning of separate channel and modality contributions.

3.4. LEVELS OF ANALYSIS: FROM MENTAL STATES TO PHYSICAL PARAMETERS AND BACK

In the analysis of language data, the same phenomena may be described at various levels, which range from the level of their physical properties (sounds of speech or written characters) to that of their representations in the mind. HIRST et al. (2000) and JASSEM (2002) point to the importance of distinguishing levels of analysis in the study of intonation. Although their approaches vary slightly, they share the intuition that one should not mix phenomena coming from different levels. For example, the acoustic level (where the acoustic signal is considered without any reference to the fact that it is speech) dramatically differs from the phonetic level (where the same phenomena can be described not only in terms of acoustic parameters but also in terms of, for example, articulatory features) and from the phonological level (where abstract entities are taken into account). This approach should definitely be employed in the field of paralinguistic analysis. Obviously, this would not involve phonology (at least in its present state) but would help to distinguish features of the signal itself, including visual signals in the case of gestural or facial display, from an auditory or visual percept. Approaching the problem from the opposite direction, one can assume that all of the paralinguistic features are “emanations” of certain mental states (LISCOMBE (2007) introduces the notion of “paralinguistic state” but he does not define it precisely). Often the mental states themselves, such as emotions or feelings, are attributed directly to speech signals or gestures: one can hear sad voices or see surprised facial expressions. But because the perception of sadness in the voice may be biased by language, culture or even individual experience, using such labels in the description of vocal or gestural display can be risky.

3.5. FROM LOCAL TO GLOBAL FEATURES

There is usually a relatively clear time window for the “linguistic” features, which is determined by the size of the linguistic units (e.g., syllables, words, phrases). Acoustic or kinetic measurements that correspond to paralinguistic features can be taken for very short stretches of speech, such as a single phone, a syllable or its core, or for an instance of gesticulation, such as a single movement of the thumb, which does not necessarily correspond to any realisations of linguistic units. However, they can also be measured for very long portions of discourse, such as entire dialogue turns and paratones. While different lengths of time windows may require different techniques and methods, as well as different interpretations of results, there are basically no restrictions on the duration of the units under analysis. Accordingly, it is evident that the range of measurement, which also refers to judgments based on perceptions, should be clearly defined and motivated, and that the interpretation of the results should also be proposed with reference to the range(s) of measurements. The labels “global” and “local” will have to remain arbitrary, both in the case of instrumental measurements and those based on perception.

4. CONCLUSION

Even though the problem of defining language and its boundaries may appear one of largely theoretical importance, it is relevant to the transcription, annotation and interpretation of interpersonal communication data, including speech and other channels. The question arises of what can and cannot be described in terms of contemporary linguistics. How can we treat the latter phenomena and, if we accept the concept, how can we understand their relationship to language proper? In the present text, an attempt has been made to suggest some directions to follow in the description and analysis of paralinguistic features. This has mostly been inspired by prosody, voice quality and gesture studies. Based on the ideas coming from these three domains, five major recommendations are made for the description and analysis of the paralinguistic component of language: to account for the multi-functionality of the paralinguistic component, to cover its multimodal nature, to carefully distinguish various levels of annotation and various scopes of features, and to prefer continua above discrete, categorical annotations. Although the primary aim of the present work is intentionally limited to some practical recommendations for the description of spoken language data, a number of fundamental problems have been touched upon in the text. While they cannot be solved here, they should certainly be borne in mind by empirically oriented linguists. One of these problems is the frequent assumption that language is the pivotal element in communication, which may not always be fully justified. Secondly, the policy of rejecting data which do not fit the linguistic framework of analysis may obviously produce misleading results and make their appropriate interpretation impossible. Finally, viewing the phenomena under discussion solely through the prism of traditional linguistic methodology may distort the data or render them less precise. In this way the rich informational load they carry will be lost.

The postulates proposed above have already been implemented in work on a comprehensive annotation system for the paralinguistic component of communication. The next steps will involve testing, adjusting and formulating them in a more detailed way. Similarly, as the fundamental question of language boundaries has proven to be very relevant in low-level work with language data, the exploration of raw data obtained from natural communication events may richly inform the big linguistic theories.

ACKNOWLEDGEMENT

(1) This work is supported from the financial resources for science in the years 2010–2012 by the National Centre for Research and Development as a development project O R00 0170 12.

(2) I would like to express my gratitude to Prof. Nicole Nau for her insightful comments on this text.

REFERENCES

- ABE Isamu. 1967. "On Bio-phonetics." *Bulletin of the Tokyo Institute of Technology* 79, 53–57 (cited from CRYSTAL 1976).
- ABELIN Ása, ALLWOOD Jens. 2000. "Cross Linguistic Interpretation of Emotional Prosody." *SpeechEmotion-2000*, Newcastle, 110–113.
- ABERCROMBIE David. 1967. *Elements of General Phonetics*. Chicago: Aldine.
- ALAHVERDZHEVA Katya, FLICKINGER Dan, LASCARIDES Alex. 2012. *Multimodal Grammar Implementation, Proceedings of NAACL-HLT*, Montpellier.
- ARBIB Michael A. 2005. "From Monkey-like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics." *Behavioral and Brain Sciences* 28, 105–124.
- BANSE R., SCHERER Klaus R. 1996. "Acoustic Profiles in Vocal Emotion Expression." *Journal of Personality and Social Psychology* 70(3), 614–636.
- BÄNZIGER Tanja, SCHERER Klaus R. 2005. "The Role of Intonation in Emotional Expressions." *Speech Communication* 46, 252–267.
- BASSO Keith, SELBY Henry (eds.). 1976. *Meaning in Anthropology*. Albuquerque: University of New Mexico Press.
- BAVELAS Janet. 1994. "Gestures as Part of Speech: Methodological Implications." *Research on Language & Social Interaction* 27(3).
- BENTHALL J., POLHEMUS T. (eds.). 1975. *The Body as a Medium of Expression*. London: Institute of Contemporary Arts.
- BIEMANS Monique. 2000. *Gender Variation in Voice Quality*. PhD Thesis, Utrecht: LOT.
- BOLSHAKOV Igor, GELBUKH Alexander. 2004. *Computational Linguistics: Models, Resources, Applications*. Mexico: IPN-UNAM-FCE.
- BRAUN A., MASTHOFF H.R. (eds.). 2002. *Phonetics and its Applications. Festschrift for Jens-Peter Koester*. Wiesbaden: Franz Steiner Verlag, 289–297.
- BROWN George. 1990. *Listening to Spoken English*. London: Longman.
- BROWN Keith (ed.). 2006. *Encyclopedia of Language and Linguistics*. Amsterdam–Boston: Elsevier, vol. 5.
- BUNT Harry. 2011. "Multifunctionality in Dialogue." *Computer, Speech and Language* 25, 225–245.
- BUNT Harry, ALEXANDERSSON Jan, CARLETTA Jean, CHOE Jae-Woong, CHENGYU FANG Alex, HASIDA Koiti, LEE Kyo-ong, PETUKHOVA Volha, POPESCU-BELIS Andrei, ROMARY Laurent, SORIA Claudia, TRAUM David. 2010. "Towards an ISO Standard for Dialogue Act Annotation." In: *Proceedings of LREC 2010, the Seventh International Conference on Language Resources and Evaluation*, Malta, May 16th – 23rd 2010.
- BURKHARDT Franz, AUDIBERT N., MALATESTA L., TUERK Oytun, ARSLAN L., AUBERGE Veronique. 2006. "Emotional Prosody – Does Culture Make a Difference?" In: *Proceedings of Speech Prosody 2006 Conference*, Dresden.
- CAMPBELL Nick. 2002. "Towards a Grammar of Spoken Language: Incorporating Paralinguistic Information." In: *Proceedings ICSLP 2002*, Denver, Colorado.
- CAMPBELL Nick, MOKHTARI Parham. 2003. "Voice Quality: The 4th Prosodic Dimension." In: *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS'03)*, Barcelona, Spain, 2417–2420.
- CAREY John. 1980. "Paralanguage in Computer-mediated Communication." In: *Proceedings of the 18th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg PA.
- CASSEL Justine, STONE Mathew, DOUVILLE Brett, PREVOST Scott, ACHORN Brett, STEEDMAN Mark. 1994. "Modeling the Interaction Between Speech and Gesture." In: *Proceedings of Cognitive Science Society Annual Meeting 1994*, George Institute of Technology.
- CHEN Aoju. 2005. *Universal and Language-specific Perception of Paralinguistic Intonational Meaning*. PhD Thesis, Radboud Universiteit Nijmegen.
- CHEN Aoju, GUSSENHOVEN Carlos, RIETVELD Toni. 2004. "Language-Specificity in the Perception of Paralinguistic Intonational Meaning." *Language and Speech* 47(4), 311–349.
- CHEVRIE-MULLER Claude, SEGUIER N., SPIRA A., DORDAIN M. 1978. "Recognition of Psychiatric Disorders From Voice Quality." *Language and Speech* 21(1), 87–111.
- CHOMSKY Noam. 1993. "A Minimalist Program for Linguistic Theory." In: HALE & KEYSER 1993: 1–52.
- CLARK Alexander, FOX Chris, LAPPIN Shalom (eds.). 2010. *The Handbook of Computational Linguistics and Natural Language Processing*. Malden–Oxford: Wiley-Blackwell.
- COHEN Philip R., MORGAN Jerry, POLLACK Martha E. (eds.). 1990. *Intentions in Communication*. Cambridge: MIT Press.

- CORNISH Francis. 2005. "Prosody, Discourse Deixis and Anaphora." In: *Proceedings of DPI-2005 Discourse-Prosody Interface, l'Université de Provence (Aix-Marseille I), Aix-en-Provence, from 8th to 9th September 2005*.
- CRUTTENDEN Alan. 1986. *Intonation*. Cambridge: CUP.
- CRYSTAL David. 1966. "The Linguistic Status of Prosodic and Paralinguistic Features." In: *Proceedings of the University of Newcastle-upon Tyne Philosophical Society* 1(8), 93–108.
- CRYSTAL David. 1969. *Prosodic Systems and Intonation in English*. Cambridge: CUP.
- CRYSTAL David. 1974. "Paralinguistics." In: SEBEOK 1974: 265–295.
- CRYSTAL David. 1975. "Paralinguistics." In: BENTHALL & POLHEMUS 1975: 162–174.
- CRYSTAL David. 1976. "Paralinguistic Behaviour as Continuity Between Animal and Human Communication." In: McCORMACK & WURM 1976: 13–27.
- CRYSTAL David. 1987. *Cambridge Encyclopedia of Language*. Cambridge: CUP.
- CRYSTAL David, QUIRK Randolph. 1964. *Systems of Prosodic and Paralinguistic Features in English*. The Hague, Netherlands: Mouton.
- CUMMINS Fred & PORT Robert F. 1996. "Rhythmic Commonalities Between Hand Gestures and Speech." In: *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum Associates, 415–419.
- CUTLER Anne. 1977. "The Context-dependence of 'Intonational Meanings'." In: *Papers from the 13th Regional Meeting of the Chicago Linguistic Society*, 104–115.
- DAVIES Alan, WIDDOWSON Henry. 1974. "Reading and Writing." *Techniques in Applied Linguistics*. London: OUP, 154–201.
- DEMENKO G., JASSEM K., SZPAKOWICZ S. (eds.). 2009. *Speech and Language Technology* 11. Poznań: Polish Phonetic Association.
- DIJKSTRA Ton, DE SMEDT Konrad. 1996. *Computational Psycholinguistics*. London: Taylor and Francis.
- DURANTI Alessandro, GOODWIN Charles (eds.). 1992. *Rethinking Context: Language as an Interactive Phenomenon*. Cambridge: CUP.
- EKMANN Paul, FRIESEN Wallace. 1969. "The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding." *Semiotica* 1, 49–98.
- FOX Anthony. 2000. *Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals*. Oxford: OUP.
- FRICKE Ellen. 2012. *Grammatik multimodal. Wie Wörter und Gesten zusammenwirken*. De Gruyter Mouton.
- FRICKE Ellen, LAUSBERG Hedda, LIEBAL Katja, MÜLLER Cornelia (in prep.). *Towards a Grammar of Gesture: Evolution, Brain, and Linguistic Structures*. Series: Gesture Studies. John Benjamins.
- GARROD Simon, PICKERING Martin. 2000. *Language Processing*. London: Psychology Press.
- GHIÒ Alain, WEISZ F., BARACCA G., CANTARELLA G., ROBERT D., WOISARD V., FUSSI F., GIOVANNI A. 2011. "Is the Perception of Voice Quality Language-dependant? A Comparison of French and Italian Listeners and Dysphonic Speakers." In: *Proceedings of InterSpeech 2011*, Florence, Italy, 525–528.
- GIBBON Dafydd. 2011. "Modelling Gesture as Speech." *Poznań Studies in Contemporary Linguistics* 47(3), 447.
- GOBL Christer, CHASAIDE Ailbe Ní. 2003. "The Role of Voice Quality in Communicating Emotion, Mood and Attitude." *Speech Communication* 40, 189–212.
- GUMPERZ John. 1992. "Contextualization and Understanding." In: DURANTI & GOODWIN 1992: 229–252.
- GUMPERZ John. 1995. "Mutual Inferencing in Conversation." In: MARKOVÁ et al. 1995: 101–123.
- GUSSENHOVEN Carlos. 2002. "Intonation and Interpretation: Phonetics and Phonology." In: *Proceedings of Speech Prosody 2002 Conference*, Aix-en-Provence.
- GUSSENHOVEN Carlos. 2004. *Phonology of Tone and Intonation*. Cambridge: CUP.
- HALE Kenneth Locke, KEYSER SAMUEL JAY (eds.). 1993. *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*. Cambridge, MA: MIT Press.
- HART Johan, COLLIER Rene, COHEN Antonie. 1990. *A Perceptual Study of Intonation*. Cambridge: CUP.
- HAUSER Marc D., CHOMSKY Noam, FITCH William T. 2002. "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" *Science* 298, 1569–1579.
- HEUSINGER K. von. 1999. *Intonation and Information Structure*. Habilitationsschrift, accepted by the Faculty of Philosophy, University of Konstanz.
- HILL Archibald E. 1952. "A Note on Primitive Languages." *International Journal of American Linguistics* 18, 172–177.
- HILL Archibald E. 1958. *Introduction to Linguistic Structures*. New York: Harcourt, Brace and World.

- HINDE Robert Aubrey (ed.). 1972. *Non-verbal Communication*. London: CUP.
- HIRSCHBERG Julia. 2002. "Pragmatics of Intonational Meaning." In: *Proceedings of Speech Prosody 2002 Conference*, Aix-en-Provence.
- HIRSCHBERG Julia, PIERREHUMBERT Janet. 1986. "Intonational Structuring of Discourse." In: *Proceedings of the 24th Meeting of the Association for Computational Linguistics*, 136–144.
- HIRST Daniel, DI CRISTO Albert, ESPESSER Robert 2000. "Levels of Representation and Levels of Analysis for Intonation." In: HORNE 2000.
- HORNE Merle (ed.). 2000. *Prosody: Theory and Experiment*. Dordrecht: Kluwer.
- HUGGINS A.W.F. 1979. "Some Effects on Intelligibility of Inappropriate Temporal Relations within Speech Units." In: *Proceedings of the 9th ICPPhS*, vol. II, 275–282.
- HYMES Dell. 1974. *Foundations of Sociolinguistics: An Ethnographic Approach*. Philadelphia: University of Pennsylvania Press.
- ISHI Carlos Toshinori, ISHIGURO H., HAGITA N. 2006. "Using Prosodic and Voice Quality Features for Paralinguistic Information Extraction." In: *Proceedings of Speech Prosody Conference 2006*, Dresden.
- JAKOBSON Roman, WAUGH Linda. 1979. *The Sound Shape of Language*. Bloomington: Indiana University Press.
- JARMOŁOWICZ-NOWIKOW Ewa, KARPIŃSKI Maciej. 2011. "Communicative Intentions behind Pointing Gestures in Task-oriented Dialogues." In: *Proceedings of GESPIN 2011 Conference*, Bielefeld.
- JARVELLA Robert J., KLEIN Wolfgang (eds.). 1982. *Speech, Place, and Action*. Chichester: John Wiley and Sons.
- JASSEM Wiktor. 2002. "Classification and Organisation of Data in Intonation Research". In: BRAUN & MASTHOFF 2002: 289–297.
- KARPIŃSKI Maciej. 2006. *Struktura i intonacja polskiego dialogu zadaniowego. (Structure and Intonation of Polish Task-oriented Dialogues)*. Poznań: Wydawnictwo Naukowe UAM.
- KARPIŃSKI Maciej. 2007. "Selected Quasi-lexical and Non-lexical Units in Polish Task-oriented Dialogues." *Archives of Acoustics* 32(1), 51–65.
- KARPIŃSKI Maciej, JARMOŁOWICZ-NOWIKOW Ewa, MALISZ Zofia. 2009. "Aspects of Gestural and Prosodic Structure of Multimodal Utterances in Polish Task-oriented Dialogues." In: DEMENKO et al. 2009.
- KARPIŃSKI Maciej, JARMOŁOWICZ-NOWIKOW Ewa. 2010. "Prosodic and Gestural Features of Phrase-internal Disfluencies in Polish Spontaneous Utterances." In: *Proceedings of Speech Prosody 2010 Conference*, Chicago.
- KEATING Patricia A., ESPOSITO Christina. 2006. "Linguistic Voice Quality." *UCLA Working Papers in Phonetics* 105, 85–91.
- KELLER Eric. 2004. "The Analysis of Voice Quality in Speech Processing." In: *Nonlinear Speech Processing: Algorithms and Analysis, 12th–17th September, 2004, Vietri*. Berlin: Springer Verlag.
- KENDON Adam. 2005. *Gesture. Visible Action as Utterance*. Cambridge: CUP.
- KLESSA Katarzyna, KARPIŃSKI Maciej. 2012. "Annotating Paralinguistic Features in Quasi-spontaneous Speech. Adding the 'Vision' Component." In: *Abstracts of Vision and Language Workshop*. Sheffield.
- KREIMAN Jody, VANLANCKER-SIDTIS Diana, GERRATT Bruce. 2004. "Defining and Measuring Voice Quality." In: *Proceedings From Sound to Sense, June 11th–June 13th*, MIT.
- LAMEL L.F., GAUVIN J.-L. 1993. "Identifying Non-linguistic Speech Features." In: *Proceedings of EuroSpeech Conference*, Berlin, September 1993, 121–124.
- LAVÉ John. 1980. *The Phonetic Description of Voice Quality*. Cambridge: CUP.
- LAVÉ John. 1994. *Principles of Phonetics*. Cambridge: CUP.
- LAVÉ John. 1995. "The Phonetic Description of Paralinguistic Phenomena." In: *Proceedings of the XIIIth ICPPhS*, Stockholm, Sweden.
- LAVÉ John. 2009. *The Description of Voice Quality in General Phonetic*. Cambridge: CUP.
- LAVÉ John, TRUDGILL Peter. 1979. "Phonetic and Linguistic Markers in Speech." In: SCHERER & GILES 1979: 1–32.
- LIEBERMAN Philip, MICHAELS S.B. 1962. "Some Aspects of the Fundamental Frequency and Envelope Amplitude as Related to Emotional Content of Speech". *Journal of the Acoustical Society of America* 34, 922–927.
- LISCOMBE Jackson. 2007. *Prosody and Speaker State: Paralinguistics, Pragmatics, and Proficiency*. PhD Thesis, Columbia University.
- LUGGER Marko, YANG Bin. 2008. "Psychologically Motivated Multi-Stage Emotion Classification Exploiting Voice Quality Features." In: MIHELIĆ & ŽIBERT 2008: 395–410.
- MARKOVÁ Ivana, GRAUMANN Carl F., FOPPA Klaus (eds.). 1995. *Mutualities in Dialogue*. Cambridge: CUP.
- MCCORMACK W.C., WURM S.A. (eds.). 1976. *Language and Man: Anthropological Issues*. The Hague: Mouton.
- MCCNEILL David. 1985. "So You Think Gestures Are Nonverbal?" *Psychological Review* 92(3), 350–371.

- MCNEILL David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- MCNEILL David. 2006. "Gesture and Communication." In: BROWN 2006: 58–65.
- MCNEILL David, LEVY Elene. 1982. "Conceptual Representations in Language Activity and Gesture." In: JARVELLA & KLEIN 1982: 271–295.
- MCNEILL David, BERTENTHAL Bennett, COLE Jennifer, GALLAGHER Shaun. 2005. "Gesture-first, but Nogestures? Commentary on Michael Arbib 'From Monkey-like Action Recognition to Human Language: An Evolutionary Framework Forneurolinguistics'." *Behavioral and Brain Sciences* 28, 105–124.
- MIHELIC France, ŽIBERT Janes (eds.). 2008. *Speech Recognition, Technologies and Applications*. Vienna: I-Tech.
- MOZZICONACCI Sylvia J.L. 1995. "Pitch Variations and Emotion in Speech." *Proceedings of the XIIIth ICPHS*, Stockholm, Sweden, vol. 1, 178–181.
- MOZZICONACCI Sylvia J.L., HERMES Dik J. 1997. "A Study of Intonation Patterns in Speech Expressing Emotion or Attitude: Production and Perception." In: *IPO Annual Progress Report* 32, Eindhoven, the Netherlands, 154–160.
- OHALA John. 1983. "Cross-language Use of Pitch: An Ethological View." *Phonetica* 40, 1–18.
- PIERREHUMBERT Jannet, HIRSCHBERG Julia. 1990. "The Meaning of Intonational Contours in the Interpretation of Discourse." In: COHEN et al. 1990: 271–311.
- PIKE Kenneth. 1943. *Phonetics*. Ann Arbor: University of Michigan Press.
- ROACH Peter. 2000. "Techniques for the Description of Emotional Speech." In: *Proceedings of the ISCA Workshop on Speech and Emotion*, Belfast, 53–59.
- ROSSINI Nicola. 2012. *Reinterpreting Gesture as Language*. Amsterdam: IOS Press.
- SCHERER Klaus R. 2003. "Vocal Communication of Emotion: A Review of Research Paradigms." *Speech Communication* 40, 227–256.
- SCHERER Klaus R., GILES Howard (eds.). 1979. *Social Markers in Speech*. Cambridge: CUP.
- SCHERER Klaus R., BANSE Rainer, WALLBOTT Herald. G. 2001. "Emotion Inferences from Vocal Expression Correlate across Languages and Cultures." *Journal of Cross-Cultural Psychology* 32(1), 76–92.
- SCHÖTZ Susanne. 2002. "Linguistic & Paralinguistic Phonetic Variation in Speaker Recognition & Text-to-Speech Synthesis." In: *GSLT Papers: Speech Technology* 1.
- SEBEEK Thomas Albert (ed.). 1974. *Current Trends in Linguistics*, vol. 12. The Hague: Mouton.
- SILVERSTEIN Michael. 1976. "Shifters, Linguistic Categories, and Cultural Description." In: BASSO & SELBY 1976.
- SMITH Henry L. 1952. *An Outline of Metalinguistic Analysis*. Washington, D. C.: Department of State, Foreign Service Institute.
- SMITH Henry L. 1953. *The Communication Situation*. Washington, D. C.: Department of State, Foreign Service Institute.
- STURTEVANT Edgar H. 1947. *An Introduction to Linguistic Science*. New Haven: Yale University (cited from CRYSTAL 1976).
- TENCH Paul. 1990. *The Roles of Intonation in English Discourse*. Frankfurt am Main: Verlag Peter Lang.
- THOMPSON William Forde, BALKWILL Laura-Lee. 2006. "Decoding Speech Prosody in Five Languages." *Semiotica* 158(1/4), 407–424.
- THORPE William Homan. 1972. "The Comparison of Vocal Communication in Animal and Man." In: HINDE 1972. (cited from CRYSTAL 1976).
- TRAGER George L. 1958. "Paralanguage: A First Approximation." In: HYMES 1958: 1–12.
- TRAGER George L. 1960. "Taos III: Paralanguage." *Anthropological Linguistics* 2(2), 24–30.
- TRAGER George L. 1961. "The Typology of Paralanguage." *Anthropological Linguistics* 3(1), 17–21.
- TRAGER George L. 1964. *Phonetics: Glossary and Tables*. New York: Studies in Linguistics.
- TRASK Robert Lawrence. 1996. *A Dictionary of Phonetics and Phonology*. London: Routledge.
- WICHMANN Anne. 2002. "Attitudinal Intonation and the Inferential Process." In: *Proceedings of Speech Prosody 2002 Conference*, Aix-en-Provence.
- WŁODARCZAK Marcin. 2012. "Ranked Multidimensional Dialogue Act Annotation." In: *Proceedings of ESSLLI 2010: International Conference on New Directions in Logic, Language and Computation*. Berlin–Heidelberg: Springer-Verlag, 67–77.
- YASINNIK Yelena, RENWICK Margaret, SHATTUCK-HUFNAGEL Stefani. 2004. "The Timing of Speech-accompanying Gestures with Respect to Prosody." In: *From Sound to Sense*, June 11–June 13, 2004, Cambridge, MA, MIT.