THEORETICAL ESSAY

LANGUAGE AND LANGUAGING FROM A PHONETIC POINT OF VIEW

Fred CUMMINS (iD) ✉
University College Dublin

Bei WANG (iD) ✉
Beijing Institute of Technology

ABSTRACT

We differentiate between language-as-system, as exemplified by such
constructs as "English," or "Mandarin,'" and *languaging*, understood as a
rich set of affiliative and coordinative behaviours that involve speech. The
former is the more familiar term, and has been constructed in a specific
manner that is inextricably bound to literacy, writing and normative social
practices. But we argue that only the latter can inform us about what it
was that happened to the human species to so differentiate us from other
primates. To draw out this distinction, we lean on the contrast between
emic and etic approaches, introduced by Ken Pike and rooted in the
distinction between phonology and phonetics. We argue that an etic
approach to speech can reveal forms of languaging that are not
addressed by language-as-system. Joint speech is put forward as an
important form of languaging that can be thematised for study only if the
emic/etic distinction is taken seriously. Consequences for the self-
understanding of phonetics as a discipline are cautiously put forward.

RESUMO

Traçamos uma diferença entre língua-como-sistema, tal como
exemplificado por construtos como "inglês" ou "mandarim" e fenômenos
linguageiros (*languaging*), compreendidos como um conjunto rico de
comportamentos associativos e coordenativos que envolvem a fala. O

primeiro termo é mais conhecido de todos e construiu-se de modo a estar inextricavelmente ligada ao letramento, à escrita e a práticas sociais normativas. No entanto, defendemos que apenas o segundo termo pode informar-nos sobre o que aconteceu com a espécie humana de modo a que tanto nos diferenciemos dos outros primatas. Para extrair essa distinção, contamos com o contraste entre as abordagens êmicas e éticas, introduzidas por Ken Pike e baseadas na distinção entre fonologia e fonética. Argumentamos que uma abordagem ética para a fala pode revelar formas linguageiras que não são tratadas pela língua-como-sistema. A fala conjunta é, então, apresentada como uma forma linguageira importante e que pode ser tematizada apenas se a distinção êmica/ética for tomada a sério. Consequências para o auto-entendimento da fonética como disciplina são cautelosamente apresentadas.

KEYWORDS

Phonetics; Phonology; Joint Speech; Languaging; Gaze.

PALAVRAS-CHAVE

Fonética; Fonologia; Fala Conjunta; Fenômenos Linguageiros; Direção do Olhar.

INTRODUCTION

The relationship between phonetics and phonology has given rise to a broader distinction that found enthusiastic uptake in anthropology and in the social sciences more broadly. This distinction is known as the emic/etic distinction, and the suffixes are derived from the terms phoneme and phonetics. Within linguistics broadly considered, the emic side of the distinction originally refers to the domain of phonology, which studies the systematic organisation of contrasting abstract units, derived from sound, and having no intrinsic meaning. The etic side, of course is rooted in phonetics, which occupies itself with the study of the physical instantiation of speech in the movements of the speaker, the acoustic disturbance created in speaking, and, to a lesser extent, the physiology of auditory speech perception. Phonology studies language, while phonetics studies speech. In this article, we wish to revisit this very familiar partitioning, learning something from the manner in which the emic/etic distinction has fared outside of its linguistic home, and draw some consequences for how phonetics, as a field, might view its relation to phonology and, more importantly, to other disciplines.

This discursive revisiting of territory that might appear to be both familiar and secure is motivated by the recognition that language itself has been continually reconceived, as it relates to the philosophy of the mind (WITTGENSTEIN, 1953) and the person (LINELL, 2009), to the social organisation of life (MATURANA; VARELA, 1987) and to novel ways of construing language as distributed (THIBAULT, 2011), embodied (DI PAOLO; CUFFARI; DE JAEGHER, 2018), and foundational for diverse human lifeworlds (CUMMINS, 2018b). Language is, itself, a moving target as our understanding changes. One could go further and suggest that the term "language" has become overloaded, and is in need of some conceptual unpacking, if we are not to continually talk past one another. In this article we approach this by drawing a distinction between language-as-system and languaging-as-activity, and suggest that these two very different framings address two very different topics, each important, but in different ways. While the former is necessary to understand contemporary forms of communication and social order, the latter is needed if we are to understand what happened to our species that differentiated us so dramatically from other primates. In light of this distinction, the relation between phonetics and other disciplines may warrant fresh appraisal, and we will suggest that phonetics, as a discipline, might see itself as well-positioned to contribute to our understanding of other domains of human social organization beyond language, narrowly conceived.

1. EMIC VS ETIC

The emic/etic distinction was introduced, and richly elaborated upon by Kenneth Pike (1967). The basic distinction is this: "The etic viewpoint studies behavior as from outside of a particular system, and as an essential initial approach to an alien system. The emic viewpoint results from studying behavior as from inside the system" (ibid, p. 37). The principal "systems" Pike treats of are cultures and languages. An etic approach "treats all cultures or languages--or a selected group of them--at one time" (p. 37), while an emic approach is "applied to one language or culture at a time." Some of the main contrasts drawn out by the emic/etic distinction as introduced by Pike are provided in Table 1.

Etic	Emic
Directly measurable	Description or measurement only relative to other system internal units
Identification of "units" using criteria external to the system	Identification only with respect to other systemic elements
Units may lack integration	Every unit functions within a larger structural unit
Units may be identified using criteria established before studying any particular system	Units must be discovered using criteria provided by the specific system being studied
Differences between units may be established straightforwardly using measurement	Differences between units arise only when they "elicit different responses from people using the system."

Table 1. Comparison of etic and emic approaches to the interpretation of behaviour

The emic/etic distinction originates in the relation between phonetics and phonology. Pike, however, enthusiastically extended the framing to many other levels of linguistic and behavioural structure. Thus, we find a proliferation of hypothetical emic units within very different structural domains: behavioreme, syntagmeme, morpheme, tagmeme, acteme, kineme (a non-verbal acteme, contrasting with the phoneme, which is a verbal acteme), hyperphoneme, the "emic syllable" (syllable-eme seems to have been a lexical innovation too far), hypermorpheme, sememe, and more. Some of these, such as morpheme, remain terms of art. Others have disappeared from the common technical vocabulary.

Two non-linguistic examples Pike uses to tease out these ideas more broadly are noteworthy. The first considers a church service, and the second, a football game. These examples are chosen to illustrate the broad applicability of the emic approach to characterising structured behaviour. Each is sufficiently rich to admit of arbitrarily varied forms of description. Pike refers to the "indeterminacy of focus" of an analytical observer. Thus one could focus on the game as a whole, on its situation within a several month-long season, on one play within the game, on a part of that play, etc. Alternative framings are available, e.g. by adopting the perspective of a vendor, coach or referee. The church service similarly presents no single hierarchical structure for analysis, and the identification of a coherent whole, such as the service, must use emic, rather than etic, categories, if the analysis is to make sense of the domain as organised. An emic analysis typically identifies a hierarchy of units, e.g. the church service, below that the section in which announcements are made,

below that the individual announcement, sentence, word, etc. The hierarchy extends no further down than those units which an "ordinary participant" might attend to. An emic description of a church service would not extend to the itemisation of phonemes employed, which belong instead to the structured domain inhabited by the phonologist.

We note in passing that it is unenlightening to describe the etic approach as being concerned with "physical" observables, while the ontological status of emic elements will differ from case to case. Etic approaches must, of course, employ measurements that admit of unproblematic interpretation by all concerned parties, but given the complexity and richness of human social organization and behaviour, these will not, in general, reduce to measurements of variables employed in physical theory. There is no agreed interpretation of the term "physical" that can serve the many cases in which etic and emic accounts co-exist, and it is simply more helpful to cast the distinction as one of insider and outsider accounts with reference to a specific system boundary. Crucially, measurements in an etic frame are assumed to be secure and uncontested by the observing community.

This extrapolation from linguistic concerns to a much broader attempt to characterise ordered behaviour in culturally saturated domains was taken up by many others. The need to accommodate insider and outsider views of cultural forms of organisation arises, for example in cultural anthropology (HARRIS, 1976), cross-cultural psychiatry (MARANO, 1982), comparative legal studies (MORRIS *et al.,* 1999), ethnomusicology (ALVAREZ-PEREYRE; AROM, 1993), cross-cultural psychology (TRINANDIS *et al.,* 1993) and many other fields of comparative social science.

One important motivation for Pike's original distinction was to avoid irresolvable epistemological issues that arise when treating of social and cultural domains, and to show how one can adopt a strongly empirical stance to underscore the analysis of structured behaviour without being forced to adopt an evaluative position with respect to the distinctions cherished by insiders. The emic/etic distinction allows social scientists to pursue appropriately disinterested empirically based analysis of events such as church services and football matches while still respecting the distinctions that matter to concerned participants. We will bear this in mind in the discussion to follow.

We note that the vast majority of phonetic studies take as given that the empirical variables they are concerned with are informative about systemic entities such as "English," "Yoruba," or "Pali," that is they contribute to the study of abstract discrete systems usually called "languages." But in turning towards *languaging*, we hope to show that phonetics can, and should, be flexibly oriented towards many kinds of emergent systems, and towards the involvement of the voice in coordinative activities that may not be systematic, but that contribute towards the emergence of the shared human world. We thus need to first contrast *language* and *languaging*.

2. LANGUAGE AND LANGUAGING

We begin by making explicit two very different questions, each of which opens up areas of inquiry of profound concern to our understanding of human living. Each of the two questions stems from an urgent desire to understand the construction of the human lifeworld, yet the two questions do not lead to the same object of inquiry. That is, the term "language" means something very different within the frame of each of the questions, such that some differentiation of terms hitherto subsumed under a single label seems necessary. In line with some other researchers, we will propose using "language" for that object framed by the first question, and the verb "languaging" for that which appears as the target of the other. While "language" picks out a specific kind of structured domain, "languaging" will be presumed to provide a richer empirical ground from which many structured domains might be built, but we will not assume "languaging" to pick out a single system, or to be necessarily committed to a specific set of observable variables.

The first question pertains to communication. It might be phrased as "How do humans communicate symbolic content from one person to the other?" This is the frame with which language is most usually constructed. It assumes the existence of *systems* (e.g. systems called "English," "Berber" etc.), within which abstract units (phonemes, morphemes, etc.) display contrastive paradigmatic organisation, rule-based syntagmatic concatenation or structural alignment, and that such systems allow a distinction, for insiders, between well-formed and ill-formed structures. Equipped with such contrasting elements and associated rules, communication, in the sense of the transmission of encoded content from producer (speaker, writer) to consumer (listener, reader), is possible. When linguists speak of "language," it is to such systems that we typically refer, and we distinguish clearly between the linguistic aspects of any token (utterance, text) and the non- or para-linguistic aspects, based on whether the variable of interest participates in this systematic organisation. Thus font color and speech tempo are not considered linguistic variables, whereas morpheme identity, or phonemic identity are. This is familiar ground.

The second question is equally motivated by a broad concern to make sense of human living. The question is "What was it that happened to our species, *homo sapiens,* that so profoundly transformed us from one primate group among many, into a globally distributed, socially organised agent capable of drastically altering the very processes of life we live among?" Language, and the associated capacity for reason, have played outsized roles in the characterisation of our species as qualitatively different from all other animal forms. These two notional capacities, language and reason, are not independent of each other, of course, and the desire to properly grasp what it is that appears to make such an enormous difference in the trajectory of one species has been fuelled by theological, moral, and anthropological concerns that are as disputed today as ever. But in such discussion, the

meaning of the term "language" is often taken as co-extensive with the systematic view just outlined. We wish to argue that this is unhelpful, that the two questions, each ostensibly about "language," are asking very different things, that answers to the former cannot speak directly to the latter, and that a broader notion of "languaging" is therefore required. We will elaborate upon some important examples of languaging that are neglected by over-zealous use of the term "language." Our overall goal is to contribute to the self-understanding of the field of phonetics in light of this distinction.

2.1. LANGUAGE CONSIDERED AS A SYSTEM

While linguistics as a whole is gloriously diverse, two formal paradigms have dominated the study of language during the 20th Century. These are the structuralist approach whose origin is most canonically associated with the work of Ferdinand de Saussure, and the generative tradition, which is indelibly associated with Noam Chomsky. Historical partitions are never neat, and frequently misleading, but it would not be unreasonable to loosely view these as developments of the first and second halves of the century, the latter coming after the computational turn that generated information theory (SHANNON, 1948) the digital computer, and, somewhat later, the functionalist view of cognition associated with cognitive psychology, and feeding back to influence the further development of ideas within the sciences of mind and of the person more generally.

Saussure famously abstracted away from much of the complexity of any token of language (an utterance, a text, in a specific context), by seeking to characterise *langue* rather than *parole*. This is where the formal characterisation of language, viewed as a self-contained system of contrasting elements organised at different levels, first becomes explicit. Building on an idea from the 19th Century (BAUDOUIN DE COURTENAY, 1972) the phoneme became central to the new scientific domain of phonology, informing the seminal ideas of the Prague school, and the field of structural linguistics as a whole.

The idea of a language as an abstract formal system, recognisable and capable of study in its own right, continued with the computational turn and the advent of the generative programme. The Chomskian distinction between *competence* and *performance* is not the same as *langue* vs *parole* as it refers to capacities of a language user, rather than to a Platonic realm of structure, but it has the same effect of making the object which linguists study an abstraction from any specific token of language use, thus adopting an *a priori* stance that detaches language from its context of use. Here, as with structural approaches, language is cast as a medium of communication, understood as the formal means by which intended meanings are encoded, passed, and received from producers to consumers, within a code such as French or Yoruba. This is how the topic of language in introduced to students, as we can see from these quotes from the influential introductory textbook "An Introduction

to Language" by Fromkin and Rodman (and latterly Hyams) (First published in 1974, but still in current print as of 2018. Quotation is from the 6th edition of 1998): "The capacity for language, perhaps more than any other attribute, distinguishes humans from other animals … When you know a language, you can speak and be understood by others who also know that language" (FROMKIN; RODMAN, 1998, p. 3).

Here, language as communicative system, and language as a biological marker distinguishing *homo sapiens* from other species are yoked together. The current edition (2018) proceeds to unpack language as morphology, syntax, semantics, phonetics, and (only then) phonology, thereby establishing the theoretical core, before using this to approach such applied topics as historical linguistics, sociolinguistics, language acquisition, and language-and-the-brain. Interestingly, earlier editions considered the origin of language, and the comparison of animal modes of communication with human language at considerable length, but more recent editions have pared such questions back or omitted them altogether, leaving the characterisation of a system for communicative message passing as the undisputed core of "language."

2.2. PRESCRIPTION VERSUS DESCRIPTION

Everyone who has taught linguistics is aware of the confusion in the public at large between prescriptive and descriptive approaches to grammar. We insist that syntacticians are not prescriptive authorities waging war on split infinitives, that linguists are scientists, not authoritarians, and that the two senses of grammar, one concerned with prescribing rules, the other with describing use, are entirely separate. This, however, misrepresents things somewhat.

The notion of language as an abstract system (whether *langue* or *competence*) is something that can be developed only if we consider communicative practices within a highly normative social order, in which there are means such as curricula, school examinations, style guides, dictionaries, authorities and the like to help to distinguish the good (reflecting the underlying hypothetical system) from the bad (the mispronounced, the infelicitous, the distracted). In the absence of such socially constructed normative criteria, it does not make sense to speak of languages as individuated systems. The possibility of identifying a canonical language depends on having means to distinguish it from variants such as vernacular usage, dialect, sociolect and ideolect. The -lects could be multiplied further, for none of us speaks in the same manner in all contexts, and even the notion of a -lect draws implicitly on the underlying assumption of a coherent system. Any review of a verbatim close transcription of unpracticed and rough speech will suffice to make it clear that the raw data of speech do not deliver such coherence.

Much of the normativity required to establish the domain of the linguistic, and to separate it cleanly from the non-linguistic, is only possible after the advent of writing. It is with writing that the very notion of a sentence as a thing in its own right that can be analysed first becomes coherent. The fact that language, as an object of study, only becomes available with the onset of writing is a point that has been made repeatedly by media scholars such as Eric Havelock, Walter Ong, and David Olson. For example Havelock (1982, p. 7) notes that "[T]he alphabet converted the Greek spoken tongue into an artifact, thereby separating it from the speaker and making it into a `language,' that is, an object available for inspection, reflection, analysis." This availability changes the very nature of the communication. As Olson notes: "Modern prose is a specialized form of language ... in which the textual/sentence meaning may be *taken* as the intended meaning. What the sentence means is seen as being sufficiently articulated that it can be taken as an adequate representation of what the speaker means" (OLSON, 1996, p. 191). Such a separation in a purely oral context is not possible.

With writing, many aspects of language become available for considered study for the first time. Discrete units arranged in space, rather than flowing continuously in time, can be attended to, without any attendant context. Distinctions captured by writing are separated from the heavily overlaid fabric of situated speech. Categorical distinctions are well captured, while continuous dimensions of variation are left behind. As a result, theories of language have had a built-in bias towards accounting for just those features of speech and talk-in-interaction (Linell's phrase) that find overt expression in writing (LINELL, 2004). Writing and reading are skills typically acquired in the strongly normative setting of the classroom, while spoken language begins at home. A single classroom may include speakers with many accents, dialects, and rhetorical capacities, but it will tend to produce writing that looks similar no matter which child does the writing, and such writing will admit of comparison with neighbouring schools, while the shouts and hoots of the playground are largely unregulated.

Standardisation happens through a variety of means. Since 1635, the Académie Française has overtly attempted to exert a standardising force on French, while the King James Bible, the Times, and the BBC, among many others, have wielded great influence on English, without necessarily displaying the same overt intention (ELWERT, 2001). The normative institutions and practices that bring into being a standard vary greatly across societies and across time. Writing itself, its constitution, purpose and form, also exhibits immense variation both synchronically and diachronically. Prior to the innovation of the printing press, and the resulting flood of textual materials, standardisation was possible only for highly restricted languages such as those of scripture. The immense work of early philologists in aligning, comparing, and correcting multiple texts produced standard texts and a standardised notion of grammar, but did not feed back into everyday language use.

The standardising practices that shape the abstraction of "English" must necessarily differ greatly from those that constrain and shape a language with only a few thousand speakers. As Ong observes:

> "The mass languages with magnavocabularies relate to the spoken word with utmost complexity. They could not sustain themselves at all without writing and print. They need writing and print to steady them synchronically in dictionaries that record what hundreds of thousands of words mean now, but they need writing and print also to orient them diachronically, to inform them what these words used to mean and how they came to their present meanings" (ONG, 1977, p. 40).

The relation of a mass language (English, Arabic, Chinese, Spanish and French as some of the leading examples) to its many sub-species, regional and social variants, becomes hard to chart, as the normative practices that induce standardisation for, for example, Hiberno-English, or Palestinian Arabic, are less developed, less precise, and less authoritative than the abstract, deterritorialised covering standard. As the speaker base becomes smaller, so the rigours of standardisation must necessarily abate. Finally, although prescription obviously plays a great role in the standardisation process, the linguist is still faced with the givenness and non-negotiability of the messy world of actual utterances by and between socially complex and situated individuals. It is the phonetician, we would argue, who is on the front line here, playing a role that enables all other forms of spoken language analysis, and, we suggest, the analysis of other aspects of languaging, to which we will turn after a brief exercise in historical calibration.

2.3. LANGUAGING OVER TIME

We wish now to draw an analogy between the development over time of systems of writing, and the development over time of patterns of many kinds of coordinative and affiliative behaviour in which speech plays an important role (languaging). In both cases we are dealing with the dynamic emergence of forms that continually evolve and change, new structures and domains of systemic organization arising continually on the basis of that which emerged before, at all timescales. In drawing this parallel, we want to avoid the suggestion that either language or writing can be pinned to a specific point in time, before which one was absent, after which it was simply present.

If one were to try to understand how contemporary writing came into being, there are many landmarks one would have to note, from the first use of tally marks in Mesopotamia, through the innovation of the alphabet, the printing press, the development of widespread literacy, emergence of genres of literature, competition from other forms of media such as radio and television, electronic communication, and, most recently, the emoji, which is finding widespread uptake among a huge user base, cross-cutting divisions normally associated with monolithic languages. Fig. 1 allows a graphical comparison of the time course of writing as a whole, assuming a Mesopotamian starting date of approximately

3,000 BCE. The time span for which printing, and after that, widespread literacy, have been important considerations is shown, along with the very tiny span of electronically mediated communication which has massively changed the manner in which text is produced and consumed. The scale of the figure precludes including emojis. While emojis are undoubtedly an important part of the contemporary writing landscape, it would be obviously unsound to begin an analysis of writing as a whole by considering emojis, or even digital text exchange. Just as an understanding of what writing is and how it arose cannot rely on the analysis of emojis, an analogous view of languaging will make it clear that innovations associated with writing, and *ipso fortiter* with printing and widespread literacy, are not going to be relevant to understanding a species-wide change. (We might note in passing that the domain of historical linguistics, which deals with relations among individuated languages both documented and reconstructed, does not attempt to peer substantially further back than the origin of writing.)
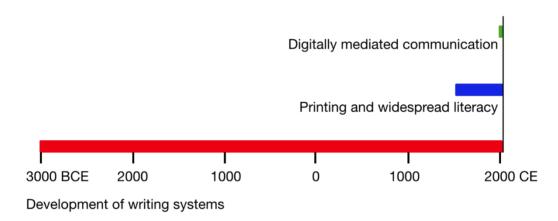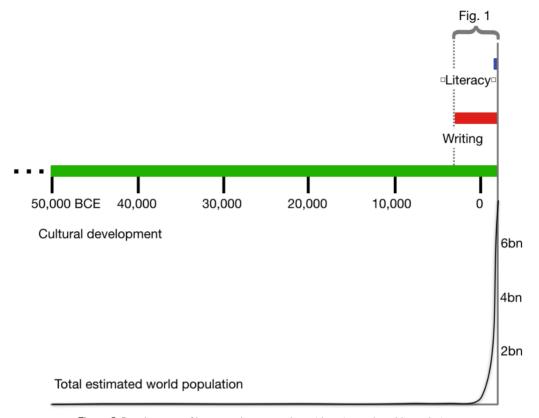

Figure 1. Timeline of the development of writing and literacy

Fig. 2 addresses the development of human cultural development in which languaging has played a role. Here, the longer time scale is a rough estimate of human cultural development, conservatively set at 50,000 years—other estimates might more than double that figure—and allows comparison with the period in which writing, and printing/literacy have held sway. The entire span of Fig. 1 is indicated for comparison. The timespan within which we find widespread literacy and an abundance of printed texts is strictly comparable in Fig. 2 to the proportion of the first figure occupied by electronically mediated communication. While exact dates are not available, there is a widely held belief that human cultural formation, evidenced by such artifacts as bone flutes, ritual burial, fire tending, etc. is presumed to be accompanied by the development of language, or as we will argue, by languaging, and we see no reason to seriously doubt this conventional assumption.

Figure 2. Development of human culture, together with estimated world population.
The span of Fig. 1 is included for comparison.

The temporal overview of Fig. 2 includes a coarse indication of how world population has changed over that period. From the plot, it is obvious that for the vast majority of the time in which *homo sapiens* exhibited the cultural development that led to contemporary language, population density was vastly less than at present, writing played no role whatsoever, and all communication was done face to face in a situation of embodied co-presence. If we are to understand what happened to our species that effected such a drastic transformation compared with other apes, it is to such situated communicative interactions that we must attend.

One might argue that this shift of perspective will reveal nothing more than proto-language, that is, that if we were to extrapolate forwards in time from such early exchanges, the principal change we would see would be the gradual emergence of domains of systematic organisation that later admit of standardisation. One might, indeed, hope to recognise the emergence of normative standards in conjunction with innovations in forms of social organisation, such as the emergence of urban settlements, or trade networks, thereby establishing a genealogy of linguistic systematicity. However, we contend that such a genealogy would be conditioned on the present form of human language as conventionally characterized and must therefore omit many of the essential features of face-to-face

interpersonal exchanges, including all and any features that do not point towards standardised and systematic language codes. Phonology—by being beholden to the idea of a coherent system of oppositions as found in modern systematised languages—is not a discipline that can contribute here, and phonetics more broadly construed may be understood to have the opportunity to help to develop a far broader notion of languaging. When we peer far back beyond the time of urban settlement, our ability to identify relevant features of human social exchange is more akin to the challenge we face in understanding vocal communication in other richly social creatures such as cetaceans (ALLEN *et al.,* 2019).

We turn now to a specific example of vocal behaviour that must, we argue, feature prominently in any account of the vocally mediated coordinative behaviours that gave rise to the complex social organization of modern humankind, and yet that is rendered invisible or inconsequential if *langue* is mistaken for languaging.

3. JOINT SPEECH AS A CORE EXAMPLE OF LANGUAGING

The present argument was originally motivated by the recognition that conventional approaches to language had failed to pick out joint speech as a distinguished form of vocal coordination of immense importance to the enactment of specific collective identities. *Joint speech refers to occasions in which multiple people utter the same words at the same time* (CUMMINS, 2013a; 2013b; 2014a; 2014b; 2018; 2019).[1] This simple empirical definition serves to draw attention to distinguished forms of human behaviour that play key roles in all human societies—in ritual and prayer, in protest, in trans-generational affiliation with sporting units, and in the education of young children, to pick out the most obvious domains. Joint speech is ubiquitous in human societies, both present and as far back as we can peer. Cummins (2018a) has argued that the oldest text identifiable as literature, the Kesh Temple Hymn which first appears in 2,600 BCE, provides strong evidence of chorusing, in that the text is liturgical, with a verse-chorus structure in which the chorus recurs in identical form after every verse. Joint speech precedes writing, appears to be ubiquitous in human cultures, and arises predictably and spontaneously among groups of pre-literate children in playgrounds. It thus does not seem to be bound to writing as the systematic domains of *langue* are.

The above definition of joint speech was introduced because no term of art existed to pick out this particular use of speech across domains. More restrictive terms exist such as choral speech, which is used both for performative recitations popular in schools, and by

---

1   A more careful definition might refer to "sounds" rather than "words," to include chant forms employing vocal productions that are not transcribable.

Kalinowski and colleages in a series of studies of means to ameliorate stuttering (KALINOW-SKI; SALTUKLAROGLU, 2003) or synchronous speech, introduced by Cummins (2002) as a laboratory task to study interpersonal vocal coordination. Specialist literatures on chanting within cultural genres, such as Gregorian chanting (APEL, 1990) or Vedic recitation (NARA-YANAN, 1994) have been developed, but the domain-transcending use of collective synchronous vocalisation in activities that found social order, enact identity, and transiently create a collective subjectivity had not been thematised in this way.

The English word "chant" is ambiguous with respect to whether the words uttered are spoken or sung, and a survey of joint speech quickly reveals that when human vocal activity is framed in this manner, no firm border between speaking and singing exists. We find examples from clearly spoken, as in the collective swearing of an oath of allegiance, to clearly sung, as in monastic liturgy, with many intermediate forms. Repetition in joint speech is the norm, rather than the exception, and this rapidly leads to rhythmic exaggeration and melodic stylisation (CUMMINS, 2019), and so a focus on joint speech may allow us to recognise commonality and continuity where a focus on linguistic systematicity may hide things by virtue of the background assumption of categorical distinctions. The familiar categorical distinction that drives a wedge between language and music can be understood differently when we take the very long view, where languaging and, we might venture, musicing (KRUE-GER, 2015; LASSITER, 2016) both appear as coordinative and affiliative activities grounded in auditory/motor sense-making.

Another obvious way in which joint speech has failed to be singled out within linguistics that focuses on abstract communicative systems is that in joint speech, the familiar distinction between speaker and listener is no longer present, or plays no role, as participants are both speakers and listeners simultaneously. As such, it does not fit a model of communicative message exchange, despite its centrality to highly valued practices that found societies and enact identities. Upon further examination we find that intelligibility is often of marginal relevance in joint speech, and that there are many examples of joint speaking in a language that is unfamiliar to those taking part. The use of Coptic or Ge'ez in religious rites, or the widespread global use of the English song "Happy Birthday" provide well-known examples.

It is also the case that the texts uttered are not the unique creative events that have played such an important role in founding a generative theory of language. As Rappaport (1990) notes, it is virtually definitional of ritual texts that they are not encoded by the performers, but are authored elsewhere.

For these, and other reasons considered in greater detail in Cummins (2018a), joint speech has been readily observable as a clear example of vocal coordination, presumably much older than writing, yet it has not contributed to developing the notion of language. Here, we suggest, the broader notion of languaging may provide us with a way to distinguish the object of study that arises as we try to understand what happened to

our species, and may help to avoid confusion with issues that arise in the study of language-as-communicative-system.

4. COMMUNICATION, COORDINATION, COMMUNION

In the introductory section, we noted that the term "language" represents something of a moving target, as competing, overlapping, and divergent interpretations of the term have developed within different research traditions, and in attempts to come to terms with many different aspects of human living. Some of these innovative traditions have made use of the term "languaging" and elaborated on it in diverse ways.

The term ``languaging'' has served several purposes in theoretical discussion (BECKER, 1991; LOVE, 2017). Languaging, as understood here, stems from the work of Umberto Maturana and Francisco Varela (1987) as a way of coming at the recursive and reciprocal nature of language behaviour, that is, languaging happens between people, as a series of concrete events, each building on the last. It is not to be understood either as a capacity of one person, or as an abstract system. Others have built on this distributed understanding of linguistic coordination, to distinguish between first-order and second-order languaging (LOVE, 2017; THIBAULT, 2017; VAN DEN HERIK, 2017). First order languaging is the situated whole-body interaction among persons, including vocal activity, but also all associated body movements, that leads to coordination of social activity. Such activity takes place at multiple timescales, and involves many kinds of processes, from local neural or bodily dynamics to the flow of whole social events (LOVE, 2017; THIBAULT, 2011). Second order language is a theoretical construction, made possible by reflecting on the concrete events of first order languaging, and producing such abstracta as "English," etc. It is noteworthy that first order languaging has been characterised as distributed among persons, rather than cast as a capacity or achievement of a single individual. This is in line with the broader notion of cognition and language as distributed phenomena (HUTCHINS, 1995; COWLEY, 2011), while the construction of language as a systemic code collapses back to the cognitivist framing of minds as individual, unobservable, and private, a position that is entirely familiar from everyday discourse, but is also epistemologically challengeable and contested (STEWART; GAPENNE; DI PAOLO, 2011; CHEMERO, 2011).

Although it is not usually expressed in these terms, we might say that first order languaging admits of description from an etic perspective, while second order language is necessarily emic in nature. Using the emic/etic distinction, rather than referring to first- and second-orders might sensitize us to the fact that the many kinds of activity, recurrences and coordinations subsumed under the label of languaging may contribute to the emergence of more than one kind of systemic organisation or structured domain. Indeed,

the fundamentally different views of cognition and mind that underlie the distributed cognition and the cognitivist position might remind us that the emic/etic distinction has found useful application in domains in which the emic stance is subscribed to by practitioners, but does not compel assent by external observers. So, for example, an etic account of the Roman Catholic sacrament of the eucharist would document liturgical form, event sequences, historical development, aesthetic qualities, etc., while an emic account would note that the event of transsubstantiation takes place at a particular moment in the ritual, after which the bread is changed in substance. Emic and etic accounts can exist in felicitous parallel as long as such borders are clear, as Pike's own use of a church service and a football match remind us.

In order to pull this overly theoretical discourse back to the applied and grounded work of phonetics, we will ask whether it might be of advantage to consider joint speech as a contributor to other forms of systemic organisation beyond language-as-system (langue) and we will suggest that there are strong grounds for including gaze among the empirical languaging variables that phoneticians might be sensitive to, as a matter of routine.

Before that, we wish first to draw out distinctions between the terms *communication, coordination,* and *communion* that can be of assistance in understanding disagreement and divergence among such approaches.[2] In order to avoid getting bogged down in the many rich issues that necessarily arise, we will limit ourselves to a few salient distinctions we believe to be probably generally acceptable, and likely to cause confusion if not recognised. We do not wish to proselytize for terminological uniformity. That would be unhelpful and likely to be rejected. We merely hope to highlight some distinctions that necessarily arise as we turn from a single account of language as communicative system, to pluralist approaches that address a wide variety of affiliative and coordinative activities that bind us together in various ways at various timescales. No attempt to be comprehensive in covering such rich territory would be plausible, and we touch only on work that is relevant to the few distinctions we wish to draw.

Although *communication* may be loosely used to describe any kind of interaction in which two or more individuals affect each other, this very broad sense obscures some important distinctions. When used in the narrower sense of encoded message passing, communication now refers to a much more constrained activity in which two or more individuals exchange information-bearing messages. Information, here, is meant in the Shannonian sense (SHANNON, 1948) and to function in this sense, it is a formal necessity that each participating system share a common vocabulary or repertoire of code units (phonemes, morphemes, words, etc.) so that messages formulated within the domain of one participant can

2   A related discussion in Jones (2016) addresses communion or common ground in the theories of Herbert H. Clark and Roy Harris.

admit of interpretation when received by another. Shannon's theory of information was developed to allow quantitative treatment of encoded signal passing. It was not developed to describe human language activities. In order for the formal language of Shannonian information to be applied in the case of human language, some commitment is necessary to characterising speakers and listeners as possessing in some form or another means for encoding and decoding and for the alignment of tokens uttered by one and received by another. Sender and receiver, or speaker and listener, must thus be comensurable and well-aligned. As long as language is abstracted away from the concrete interactions of embodied individuals, these desiderata are simply assumed.

When we consider how such an abstract system might find implementation in persons, the same desiderata can be met within some cognitivist theories of mind. Indeed, the cognitivist programme in which generative linguistics played such an important part, could be viewed as constructing just the necessary theoretical architecture to allow precisely this interpretation, within a broader conception of the person as information processor. As widespread and conventional as such views are, they rest on metaphysical and theoretical foundations that are by no means universally accepted. Theories of extended mind (CLARK; CHALMERS, 1998), embodied cognition (VARELA; THOMPSON; ROSCH, 1992; CHEMERO, 2011), and enaction (STEWART; GAMENNE; DI PAOLO, 2011; DI PAOLO; CUFFARI; DE JAEGHER, 2018) represent some of the more influential currents that have emerged in contrast to the cognitivist paradigm. In each case, the commitment to an unseen domain of mind, separate from the necessary entailments of incarnation or physical instantiation, is rejected, such that the framing of language as message passing no longer appears as a viable alternative.

An alternative to communication is provided by the generous and flexible term *coordination*. By coordination, we can understand mutuality in interaction, such that two or more individuals become, in some respects, non-independent by virtue of participation in the interaction. This term can apply to interactions that we would never be tempted to construe as linguistic, e.g. to the mating behaviour of insects, or the flocking of geese, but it provides a useful term with which to consider concrete examples of interaction among embodied individuals. To describe two (or more) individuals as coordinated is to adopt a stance that sees each as describable as an autonomous dynamical system, and to recognize that in the interaction, the total number of effective degrees of freedom of the ensemble is less than the sum of the degrees of freedom of the individuals considered separately. Less technically, if you and I engage in a handshake, then for the duration of the handshake, our two hands become transiently non-independent, each constraining the other in reciprocal dance.

This kind of characterisation of interactions leads naturally to the adoption of the (technical) language of dynamics rather than symbols (KELSO, 1995; CHEMERO, 2011), and to the consideration of specific concrete instances of interactions. The very generality of the notion

means that coordinative and dynamical accounts may address many kinds of interaction beyond those of concern to us here, and there is an explanatory gap that remains to be bridged if we are to provide a continuous account from such concrete interactions to a traditional account of systematic symbolic language use. There have been bold attempts to bridge such a gap in the treatment of language (RACZSAZEK-LEONARDI; KELSO, 2008; DI PAOLO; CUFFARI; DE JAEGHER, 2018) and in cognition more generally (FROESE; DI PAOLO, 2019), and we are reminded of the integrationist theory of Roy Harris (1998) which emphasizes the necessity for verbal signs to be continuous with non-verbal meaning-making activities of human living, thus making the isolation of an abstract informational domain strictly impossible. This integration applies not only to the suite of activities of an individual, but to the necessary common basis that multiple individuals must share in order to coordinate.

Which leads us to what we will term *communion*, a term that more likely to be found in discussing relations among churches than people, but that we wish to suggest draws out the notion of coordination leading to alignment that provides the foundation or ground from which groups may enact a common world. Without such prior alignment, communicative interaction in the strict sense is an impossibility, and such alignment may provide the basis for much more than a language system. The notion of communion suggests two (or more) systems that are in reciprocal contact and exchange with each other, sharing a great deal. The mathematician George Spencer-Brown expressed our view of communication and communion well when he noted:

> [T]he characteristic of communication is that what goes on goes on at the same level. ... [I]f there is no communion, as indeed there sometimes is not, then what is communicated, when it reaches the other end, it not understood. The more perfect the fit on the communion level, the less needs to be communicated, the more that can be crossed from one being to another in fewer actual communicated acts. (SPENCER-BROWN, 1973)

Communion, in this sense, is similar to the more familiar notion of common ground, and points also towards the venerable notion of sensus communis, which can be traced back to Aristotle. These related notions all seek to capture the fundamental idea that in order for joint meaning making to be possible, two (or more) participants must share a background, a foundation, or an embedding, and this precondition for communication is not, itself, part of any communicative exchange, but a precondition for the same. The more participants share, the less needs to be communicated from one to the other. One thinks of the economy of communication between married partners who have decades of experience in sharing a lifeworld. Under these circumstances, a great deal can be achieved with a minimum of effort. By way of contrast, one can shout at an earthworm all day without communicating anything.

Bearing in mind that most people, at most times, lived in small, widely distributed ethnic groups and all communication was face to face, one can readily see that a great deal of commonality or alignment is to be expected between those who live closely together. But it

is an open question just which activities contribute to such alignment, laying the foundation for a shared world in which communication, *sensu strictu*, is a possibility. Here, we suggest, the discipline of phonetics has an important role to play, a role that is rendered invisible if we fail to distinguish between language-as-system, and languaging as a suite of diverse activities. What are the activities that give rise to the necessary sense of communion that can facilitate coordination, leading, *inter alia*, to communication? An etic perspective need not be beholden to any one emic domain. Speech and associated activities participate in the creation of a shared world in more ways than one. Can we point to ways in which phonetics, with its strongly empirical attachments, can contribute to a more complete understanding of languaging in the broadest sense? We now lay out two routes that suggest themselves, with no attempt whatsoever at being comprehensive.

5. PHONETICS OF LANGUAGING

5.1. GAZE

The historical perspective provided above has served to remind us forcefully that languaging has almost always been an affair that has happened among individuals who are in each other's embodied presence. While the voice is undoubtedly extremely important in such situations, and the effective involvement of the voice is one reason we might describe a given activity as languaging, we suggest that the role of gaze deserves consideration as a fundamental phonetic variable of central relevant to the activity of languaging (though probably irrelevant to considerations of language). Treated in this way, gaze would belong along with analysis of fundamental frequency, spectral prominences and distributions, intensity profiles etc., in consideration of languaging as an embodied activity of co-present individuals.

One important motivator for this suggestion lies in consideration of evolutionary changes within the lineage from *homo sapiens* and extending back 5 or 6 million years to the last common ancestor of the genera *homo* (humans) and *pan* (chimpanzees and bonobos). While there was no major change in brain morphology on either branch of the evolutionary tree, there was one obvious morphological change that has been singled out as potentially significant: the change of the sclera of the eye from a dark colouration to white. A white sclera contrasts strongly with a coloured iris and dark pupil, making the eyes far more informative about the direction of gaze, and hence about the distribution of attention of conspecifics, in humans than in other great apes. Tomasello and colleagues have introduced this as the Cooperative Eye Hypothesis (2007). In elaborating upon this hypothesis, they have shown that human children follow the direction of gaze, while apes, who are also interested in what their conspecifics are looking at, rely on the much poorer signal of head direction (TOMASELO *et al.*, 2007). They have further pointed out how this small biological

change has the knock-on effect that infants and small children are immersed in situations of joint attention from an early age. This provides a possible basis for shared understanding and shared intentionality, just as a theory of languaging would require (TOMASELLO *et al.*, 2005). The cooperative eye hypothesis provides a small but important piece in the overall puzzle of how a single species might be transformed in a relatively short space of time, by taking the burden of explanation out of the realm of the biological and into patterns of interaction that lead to interpersonal coordination, or, as we might suggest, communion.

Along with gesture, gaze has been long recognized as important in the interpersonal coordination of spoken interaction (ARGYLE; COOK, 1976), and modified patterns of gaze have been robustly associated with the altered communicative patterning of people with autism (DICKERSON *et al.*, 2005). Gaze, and even blinks, have been demonstrated to be integral parts of spoken interaction in a face to face situation (CUMMINS, 2012). Reproducing plausible gaze patterns is an explicit goal in the development of embodied conversational agents (CASSELL *et al.*, 2000), while gaze is well known to bear important pragmatic functions, e.g. in disambiguation (HANNA; BRENNAN, 2007), and is an indirect index of shared belief among conversation partners (RICHARDSON; DALE; TOMLINSON, 2009). None of these observations would lead one to suggest that gaze is important phonologically, but each finding adds substance to the claim that gaze is an essential component of interpersonal coordination in face to face situations. As such, we would suggest that gaze is part and parcel of the means by which communion arises, that is prior to communication.

One could extend such argument to many other variables that co-occur in the context of embodied interpersonal coordination, including, and especially, gesture (WAGNER; MALISZ; KOPP, 2014), but we point here to gaze because of the suggested role of perceptible gaze direction in bringing into being situations of joint attention for humans, as opposed to apes, from a very young age, and thus plausibly central to any characterising of languaging.

5.2. THE PHONETICS OF JOINT SPEECH

How might one approach joint speech phonetically, and what kind of etic/emic consideration arise? Phonetic characteristics are those that are based on processes of measurement and description that are uncontested to a given group of observers, without prior commitment to the elements and structures of a systematic domain of organization. Such observations might be expected to transcend the highly structured and elaborate domains of specific forms of cultural activity in which joint speech is found. It is empirical phonetic observation that can allow us to recognise important commonality that unites the chants of football fans, Benedictine monks and school children, thus drawing to our attention these foundational activities as part of the emergent legacy of languaging. An emic perspective must, of necessity, address those elements of a situation that are judged important or

constitutive by those within a domain, and so emic approaches will, of necessity, miss such domain-spanning commonalities.

It is not self-evident how to study joint speech experimentally, and therein lies, we suggest, an important lesson for phonetics. The empirical definition employed—multiple people uttering the same words at the same time—can be reproduced without further ado in the laboratory. I have coined the term "synchronous speech" for just this case, and there is much that can be studied (CUMMINS; ROY, 2001; CUMMINS, 2002, 2003, 2004, 2007, 2009; CUMMINS; LI; WANG, 2013). Synchronisation is readily measurable, and the tightness of synchronisation can be probed under different experimental conditions, e.g. with and without visual feedback (CUMMINS, 2003). Variability across speakers is radically reduced when constrained to read in synchrony (CUMMINS, 2004), suggesting that synchronous speech, as well as being an object of study, might be a tool of interest to the phonetician in studying read speech. But synchronous speech is a rather mechanical analogue of joint speech as found in the rich culturally saturated contexts in which it arises. Adding more speakers, beyond the dyad, does not serve to alter synchronous speech in any obvious way, or to make it more closely approximate the affectively saturated form of joint speech uttered in context (CUMMINS, 2018a).

Synchronous speech has some properties that are not found in temples and schools. For example, in laboratory conditions, subjects are explicitly encouraged to synchronise with each other, which leads to a much greater degree of temporal synchrony than is found in ritual or protest. This strong experimentally induced coupling frequently leads to a very characteristic error, in which both speakers abruptly stop speaking together, e.g. after uncertainty caused by a more minor error on the part of one speaker (CUMMINS, 2012). Such abrupt cessation never happens when a speaker speaks alone, and, importantly, it also does not occur in group recitation of prayers, chants, etc. It is thus an artifact of the experimental situation.

Conversely, joint speech could be argued to have many properties that are not transportable to the laboratory. The passion of the protester, the piety of those who pray, or simply the exuberance of the supporters or children, all resist transplantation to the constrained environment of the phonetics lab. Each of these examples has its own domain-specific character, and pertinent observations that will shed light on the form of coordination involved may go far beyond the set of phonetic variables available. The ceremonial robes of a priest and the ritualised gestures of the congregation are surely relevant when studying religious ritual, but are foreign to the school playground.

When we, as phoneticians, employ standard methods designed to elucidate phonological concerns, as, for example, when we have subjects read phonemically contrasting words in a standardised carrier context ("*I said heed again; I said who'd again …*")*,* the methods employed are not theoretically neutral. Rather, they are customised to suit the emic domain

being studied. In developing materials, the structure of the phonological domain is assumed. It is not discovered (although a failure to bring forth the expected elements may lead to a revision of the associated phonological theory).

Phonetic considerations can alert us to commonalities across domains evident in observable shared characteristics. Emic considerations will be required to tease out the relevant voice (and gaze, gesture, etc.) features that bring about those domain-specific elements that matter to insiders, for whom the structure of the emic domain is meaningful. This suggests that the study of joint speech as it is found in ritual contexts will probably need to employ different methods from the study of joint speech in the context of football supporters, the joint speech of bullying calls of children in the playground may demand different methods than the study of oaths of allegiance. But this richness can serve to alert us to the delicate back-and-forth between theory and observation, between emic and etic concerns. In as much as phonetics aspires to provide the etic observations that can inform theory, theory will suggest appropriate methods and constraints that can shed light on domain-specific phenomena.

6. CONCLUSIONS

We have argued that two fundamentally different ways of beginning any inquiry into language/languaging need to be distinguished. Consideration of systemic entities like "English" sheds light on contemporary communicative practices, but cannot, we argue, contribute much to the important question of just what happened to our species that so profoundly influenced its development and its effect on the planet. To address this, we need to consider a broad range of affiliative and coordinative practices, many of them crucially involving vocal coordination, which collectively provide the necessary common ground, or communion, that is prior to communication. Phonetics, we believe, has an important role to play here, precisely because its strongly empirical stance makes it capable of providing measurement and description appropriate to the characterisation of many important activities that are foundational for shared cultural worlds.

The emic/etic distinction has proved to be of great value in allowing a principled separation between the structural domain of phonology and the empirical world of phonetics. Phonetic study has always aspired to providing expertise on its own terms, without necessary commitment to any particular theory of language. This on its own is not particularly novel. We might note in particular the vital contribution of phonetic knowledge to the clinical understanding and treatment of speech disorders, where theoretical considerations may be of marginal relevance, phonetic input to sociolinguistics (which struggles more than most

with the hegemony of *langue*) or phonetic insights to vocal behaviour such as singing. Phonetic insights have contributed to many disciplines, not all organized by *langue*.

The emic/etic distinction introduced by Pike was intended to be of use in a much broader science of human behaviour (a "unified theory" in Pike's terms). Here, we have suggested, languaging stands in need of substantial elaboration. Recognition of joint speech as an important facet of languaging provides motivation, we believe, for phoneticians to reach out to other domains from social and cultural sciences. The identification of joint speech as an object that demands serious study as a form of languaging is, after all, based entirely on phonetic criteria. The simple definition employed to delineate the phenomenon—multiple people uttering the same words at the same time—is based entirely on observation, with the background assumption that those employing the definition already agree (common ground) on what it is to utter, and that they can jointly recognise what a word is. This is not to suggest that phoneticians have the privilege of a mythical "view from nowhere" (NAGEL, 1989). Rather it is to pull back from the theoretical commitments of any particular characterisation of *langue* to a broader perspective from which the basis for many forms of coordination can be investigated.

The emic/etic distinction continues to flourish in many fields of social science, in anthropology, in ritual studies, religious studies, and beyond. In all of these fields, there is an in-built sensitivity to the necessity of keeping the emic commitments, meaningful to participants and practitioners in their own terms, separate from the grounded etic observations that can provide the empirical foundation necessary for the development and curation of any objective account. As speakers of "English," "Italian," etc. we are all, to some extent insiders; we are all language users for whom the theoretical abstractions of phoneme, morpheme, word, clause, etc. stand in no immediate need of justification (though there are dissenting voices, e.g. Port and Leary (2005)). But we, as phoneticians, stand both within *langue* and outside it.

The need to consider forms of speech, and speech in context, other than those suited to bringing forth the elements of formal phonology has been well expressed before (WAGNER; TROUVAIN; ZIMMERER, 2015). What has received less attention, if any, is the manner in which the theoretical object of "language" has been constructed in the past, and it is worth returning to the work of Ken Pike here, as he was influential in many ways beyond introducing the emic/etic distinction. He was also the first president of the Summer Institute of Linguistics, which has contributed enormously to the received view that languages are distinct things, that can be enumerated and documented, and that can be treated as naturally occurring objects. The Summer Institute of Linguistics (SIL) is also the source of the Ethnologue database, which is routinely used to inform discussion of linguistic diversity and change, as if it were simply a scientific resource. As such, the SIL and Ethnologue have played major roles in stabilising and systematising our notion of what a language (*langue*) is. The SIL,

however, is not an objective player here. It is a faith-based organisation, specifically grounded in evangelising Protestant Christianity (STOLL, 1982). Its work is not mere documentation, but the active spread of the bible. This brings another insider domain into being that demands consideration. Within the world of SIL, and by extension Ethnologue, a language is a code that admits of translation of the Christian bible. This is one of the criteria used to demarcate the boundaries of language, and is one of the principal sources of information for lesser studied or endangered languages. The SIL has a history of camouflaging its missionary concerns by presenting itself as an impartial linguistic body whose concern is the study of language, while working hand in hand with a variety of interests to further its missionary aims.

This is not the place to embark on a socio-political analysis of the specific role of SIL and Ethnologue, though such work would be a valuable contribution to contemporary linguistics. It is adduced in this context simply to show how the construction of the abstraction of language is not, and can never, be free of socially normative influences at all levels. Rather, we have sought to emphasize the importance of curating a phonetic discipline that is not beholden to a single view of language as system. Phonetics is uniquely positioned to help develop a broader picture of languaging (including joint speech) as long as practitioners are aware of the important difference between emic and etic concerns. Worryingly, Moore and Skidmore (2019) recently found that many speech scientists contributing to INTERSPEECH-2018 consistently misused the term 'phoneme' and were unaware of the difference between phonemic and phonetic levels of description. We hope that by laying out the difference between language as system, and languaging more generally, we can contribute to an important awareness that underpins the discipline of phonetics, and encourage broader investigation of social and affiliative coordination that is sensitive to broader cultural and biological concerns.

7. ACKNOWLEDGMENTS

REFERENCES

APEL, W. *Gregorian Chant*. Indiana University Press, 1990.

ARGYLE, M.; COOK, M. *Gaze and Mutual Gaze*. Cambridge University Press, 1976.

BAUDOUIN DE COURTENAY, J. *A Baudouin de Courtenay Anthology: The Beginnings of Structural Linguistics*. Indiana University Press, 1972.

BECKER, A. L. Language and languaging. *Language and Communication*, 11(1):33–35, 1991.

CASSELL, J.; SULLIVAN, J.; CHURCHILL, E.; PREVOST, S. *Embodied Conversational Agents*. MIT press, 2000.

CHEMERO, A. *Radical Embodied Cognitive Science*. MIT press, 2011.

CLARK, A.; CHALMERS, D. The extended mind. *Analysis*, 58(1):7, 1998.

COWLEY, S. J. *Distributed Language*, John Benjamins Publishing, 2011.

CUMMINS, F. On synchronous speech. *Acoustic Research Letters Online*, 3(1):7–11, 2002. DOI https://doi.org/10.1121/1.1416672

CUMMMINS, F. Practice and performance in speech produced synchronously. *Journal of Phonetics*, 31(2):139–148, 2003. DOI https://doi.org/10.1016/S0095-4470(02)00082-7

CUMMINS, F. Synchronization among speakers reduces macroscopic temporal variability. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, pages 304–309. 2004.

CUMMINS, F. Speech synchronization: Investigating the links between perception and action in speech production. In *Proceedings of ICPhS 2007*, pages 529–532, Saarbrücken, DE. 2007.

CUMMINS, F. Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, 37(1):16–28. 2009. DOI https://doi.org/10.1016/j.wocn.2008.08.003

CUMMINS, F. Gaze and blinking in dyadic conversation: A study in coordinated behaviour among individuals. *Language and Cognitive Processes*, 27(10):1525–1549. 2012. DOI: 10.1080/01690965.2011.615220

CUMMINS, F. Joint speech: The missing link between speech and music? *Percepta—Revista de Cognição Musical*, 1(1):17–32. 2013a.

CUMMINS, F. Towards an enactive account of action: Speaking and joint speaking as exemplary domains. *Adaptive Behavior*, 13(3):178–186. 2013b. DOI: 10.1177/1059712313483144

CUMMINS, F. The remarkable unremarkableness of joint speech. In *Proceedings of the 10th International Seminar on Speech Production*, pages 73–77, Cologne, DE. 2014a.

CUMMINS, F. Voice, (inter-)subjectivity, and real time recurrent interaction. *Frontiers in Psychology*, 5(760). 2014b. DOI https://doi.org/10.3389/fpsyg.2014.00760

CUMMINS, F. Comparing joint speech and synchronous speech: What happens when we add more speakers? In *Proceedings of AICS 2018 (CEUR Workshop Proceedings)*, Vol 2259, pp.:14-20. 2018a.

CUMMINS, F. *The Ground From Which We Speak: Joint Speech and the Collective Subject*. Cambridge Scholars. 2018b.

CUMMINS, F. Music, language and languaging. In CORRÊA, A. F., editor, *Music, Speech, and Mind*. Curitiba: Brazilian Association of Cognition and Musical Arts - ABCM. 2019.

CUMMINS, F.; LI, C.; WANG, B. Coupling among speakers during synchronous speaking in English and Mandarin. *Journal of Phonetics*, 41(6):432–441. 2013. DOI: https://doi.org/10.1016/j.wocn.2013.07.001

CUMMINS, F.; ROY, D. Using synchronous speech to minimize variability in pause placement. In *Proceedings of the Institute of Acoustics*, volume 23(3), pages 201–206, Stratford-upon-Avon. 2001.

DI PAOLO, E. A.; CUFFARI, E. C.; DE JAEGHER H. *Linguistic Bodies: The Continuity between Life and Language*. MIT Press. 2018.

DICKERSON, P.; RAE, J.; STRIBLING, P.; DAUTENHAHN, K.; WERRY, I. Autistic children's co-ordination of gaze and talk: Re-examining the 'asocial' autist. In RICHARDS, K.; SEEDHOUSE, P., editors, *Applying Conversation Analysis*, pages 19–37. Springer. 2005.

ELWERT G. Societal literacy: Writing culture and development. In Olson, D. R. and Torrance, N., editors, *The Making of Literate Societies*, pages 54–67. Blackwell. 2001.

FROESE, T.; DI PAOLO, E. A. Sociality and the life-mind continuity thesis. *Phenomenology and the Cognitive Sciences*, 8(4):439. 2009. DOI https://10.1007/s11097-009-9140-8

FROMKIN, V.; RODMAN, R.; HYAMS N. *An Introduction to Language*. Cengage Learning. 2018/1974

HANNA, J. E.; BRENNAN, S. E. Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, 57(4):596–615. 2007. https://doi.org/10.1016/j.jml.2007.01.008

HARRIS, M. History and significance of the emic/etic distinction. *Annual Review of Anthropology*, 5(1):329–350. 1976.

HARRIS R. *Introduction to Integrational Linguistics*. Pergamon. 1998.

HAVELOCK, E. A. *The Literate Revolution in Greece and Its Cultural Consequences*. Princeton University Press. 1982.

HUTCHINS, E. *Cognition in the Wild*. MIT press. 1995.

JONES, P. E. 'Coordination' (Herbert H. Clark), 'integration' (Roy Harris) and the foundations of communication theory: common ground or competing visions? *Language Sciences*, 53:31–43. 2016. DOI: https://doi.org/10.1016/j.langsci.2015.07.001

KALINOWSKI, J.; SALTUKLAROGLU, T. Choral speech: the amelioration of stuttering via imitation and the mirror neuronal system. *Neuroscience & Biobehavioral Reviews*, 27(4):339–347. 2003. DOI: https://doi.org/10.1016/S0149-7634(03)00063-0

KELSO, J. A. S. *Dynamic Patterns*. MIT Press, Cambridge, MA. 1995.

KRUEGER, J. Musicing, materiality, and the emotional niche. *Action, Criticism, and Theory for Music Education*, 14(3):43. 2015.

LASSITER C. Aristotle and distributed language: capacity, matter, structure, and languaging. *Language Sciences*, 53:8–20. 2016. DOI: https://doi.org/10.1016/j.langsci.2015.05.011

LINELL, P. *The Written Language Bias in Linguistics: Its Nature, Origins and Transformations*. Routledge. 2004.

LINELL, P. *Rethinking Language, Mind, and World Dialogically: Interactional and Contextual Theories of Human Sense-Making*. IAP. 2009.

LOVE, N. On languaging and languages. *Language Sciences*, 61:113–147. 2017. DOI: https://doi.org/10.1016/j.langsci.2017.04.001

MARANO, L. Windigo psychosis: The anatomy of an emic-etic confusion. *Current Anthropology*, 23(4):385–412. 1982.

MATURANA, H. R.; VARELA, F. J. *The Tree of Knowledge: The Biological Roots of Human Understanding*. New Science Library/Shambhala Publications. 1987.

MOORE, R. K.; SKIDMORE, L. On the use/misuse of the term 'phoneme'. *Proc. Interspeech 2019*, pages 2340–2344. 2019.

MORRIS, M. W.; LEUNG, K.; AMES, D.; LICKEL, B. Views from inside and outside: Integrating emic and etic insights about culture and justice judgment. *Academy of Management Review*, 24(4):781–796. 1999.

NAGEL, T. *The View From Nowhere*. Oxford University Press. 1989.

NARAYANAN, V. *The Vernacular Veda: Revelation, Recitation, and Ritual*. University of South Carolina Press. 1994.

OLSON, D. R. *The World on Paper: The Conceptual and Cognitive Implications of Writing and Reading*. Cambridge University Press. 1996.

ONG, W. J. *Interfaces of the Word: Studies in the Evolution of Consciousness and Culture*. Cornell University Press. 1977.

PIKE, K. L. *Language in Relation to a Unified Theory of the Structure of Human Behavior*. Mouton & Co. 1967.

PORT, R. F.; LEARY, A. P. Against formal phonology. *Language*, 81(4):927–964. 2005. DOI: https://doi.org/10.1353/lan.2005.0195

RACZASZEK-LEONARDI, J.; KELSO, J. S. Reconciling symbolic and dynamic aspects of language: Toward a dynamic psycholinguistics. *New Ideas in Psychology*, 26(2):193–207. 2008. DOI: https://doi.org/10.1016/j.newideapsych.2007.07.003

RICHARDSON D. C.; DALE, R.; TOMLINSON, J. M. Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science*, 33(8):1468–1482. 2009. DOI: 10.1111/j.1551-6709.2009.01057.x

SHANNON, C. E. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379– 423. 1948.

SPENCER-BROWN, G. *Transcript of the 1973 AUM Conference*. Available online at https://www.kurtvonmeier.com/aum-transcript-3. 1973.

STEWART, J.; GAPENNE, O.; DI PAOLO, E. *Enaction: Toward a New Paradigm for Cognitive Science*. MIT Press. 2011.

STOLL, D. The summer institute of linguistics and indigenous movements. *Latin American Perspectives*, 9(2):84–99. 1982.

THIBAULT, P. J. First-order languaging dynamics and second-order language: the distributed language view. *Ecological Psychology*, 23(3):210–245. 2011. DOI: https://doi.org/10.1080/10407413.2011.591274

THIBAULT, P. J. The reflexivity of human languaging and Nigel Love's two orders of language. *Language Sciences*, 61:74–85. 2017. DOI: https://doi.org/10.1016/j.langsci.2016.09.014

TOMASELLO, M.; CARPENTER, M.; CALL, J.; BEHNE, T.; MOLL, H. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5):675–691. 2005.

TOMASELLO, M.; HARE, B.; LEHMANN, H.; CALL, J. Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *Journal of Human Evolution*, 52(3):314–320. 2007. DOI: https://doi.org/10.1016/j.jhevol.2006.10.001

TRIANDIS, H. C.; MC CUSKER, C.; BETANCOURT, H.; IWAO, S.; LEUNG, K.; SALAZAR, J. M.; SETIADI, B.; SINHA, J. B.; TOUZARD, H.; ZALESKI, Z. An etic-emic analysis of individualism and collectivism. *Journal of Cross-Cultural Psychology*, 24(3):366–383. 1993.

VAN DEN HERIK, J. C. Linguistic know-how and the orders of language. *Language Sciences*, 61:17–27. 2017. DOI: https://doi.org/10.1016/j.langsci.2016.09.009

VARELA, F. J.; ROSCH, E.; THOMPSON, E. *The Embodied Mind: Cognitive Science and Human Experience*. MIT press. 1991.

WAGNER, P.; MALISZ, Z.; KOPP, S. Gesture and speech in interaction: An overview. *Speech Communication*, 57 (Special Iss.). 2014. DOI: https://doi.org/10.1016/j.specom.2013.09.008

WAGNER, P.; TROUVAIN, J.; ZIMMERER, F. In defense of stylistic diversity in speech research. *Journal of Phonetics*, 48:1–12. 2015. DOI: https://doi.org/10.1016/j.wocn.2014.11.001

WITTGENSTEIN, L. *Philosophical Investigations. Philosophische Untersuchungen*. Macmillan. 1953.