

UBCWPL

University of British Columbia
Working Papers in Linguistics

-Papers for the Interlocution Workshop-
Interlocution:
Linguistic structure and human interaction



Edited by:
Anita Szakay, Connor Mayer, Beth Rogers, Bryan Gick and Joel Dunham

July 2009

Volume 24

-Papers for the Interlocution Workshop-
Interlocution:
Linguistic structure and human interaction

Vancouver, British Columbia, Canada
May 15-17, 2009

Hosted by:
The Department of Linguistics at the University of British Columbia

Edited by:
Anita Szakay, Connor Mayer, Beth Rogers, Bryan Gick and Joel Dunham

The University of British Columbia Working Papers in Linguistics
Volume 24

July 2009

UBCWPL is published by the graduate students of the University of British Columbia. We feature current research on language and linguistics by students and faculty of the department, and we are the regular publishers of two conference proceedings: the Workshop on Structure and Constituency in Languages of the Americas (WSCLA) and the International Conference on Salish and Neighbouring Languages (ICSNL).

If you have any comments or suggestions, or would like to place orders, please contact :

UBCWPL Editors
Department of Linguistics
Totem Field Studios
2613 West Mall
V6T 1Z4
Tel: 604 822 4256
Fax 604 822 9687
E-mail: <ubcwpl@gmail.com>

Since articles in UBCWPL are works in progress, their publication elsewhere is not precluded. All rights remain with the authors.

Cover artwork by Lester Ned Jr.

Contact: Ancestral Native Art Creations

10704 #9 Highway

Compt. 376

Rosedale, BC V0X 1X0

Phone: (604) 793-5306

Fax: (604) 794-3217

Email: ldouglas@uniserve.com

TABLE OF CONTENTS

Preface.....	v
BRYAN GICK	1
Interlocution: an overview	
FREDERICK J. NEWMAYER	9
Cognitive and communicative factors in language evolution	
MOHAN MATTHEN	18
Epistemic affordances: the case of colour	
MOHINISH SHUKLA	26
Domain specificity and statistical computations in segmenting fluent speech	
DIANA ARCHANGELI	36
Public and private patterns in language	
ARLENE S. WALKER-ANDREWS.....	44
Development of communication in a social/emotional context	
OLIVIER PASCALIS AND LESLEY UTTLEY	50
Face and speech: how and when do infants understand ethnicity?	
CASEY O'CALLAGHAN	57
Is speech special?	
NAVIN VISWANATHAN.....	65
Towards embedded and embodied accounts of language use: insights from an ecological perspective	

ALESSANDRO D'AUSILIO.....	72
Motor contribution to speech perception	
SONYA BIRD	77
When socio-linguistic interaction breaks down	

PREFACE

Volume 24 of the University of British Columbia Working Papers in Linguistics (UBCWPL) series presents the Proceedings of the first ever Interlocution Workshop, which was held at the University of British Columbia in Vancouver, British Columbia, Canada, May 15-17, 2009.

We would first like to thank the authors for their submissions.

Enjoy the volume!

Anita Szakay
Connor Mayer
Beth Rogers
Bryan Gick

Interlocution: An overview

Bryan Gick
University of British Columbia

During much of the last century, linguists have described human language as a system of abstract cognitive structures, largely unconnected to other aspects of cognition or social interaction. This view has developed within a broader context of psychological models focusing on internal information processing and construction of abstract representations. Linguists now hold a deep understanding of these complex patterns, but lack satisfying explanation as to their relation to other human behavior. The Interlocution framework presented here draws together the many diverse strands of research, most of which have emerged in the past decade, that relate human communication with human interaction.

1 Interlocution

Nearly all of the approaches to linguistic theory adopted in the last half-century have focused on abstract mental representations, and may be seen as belonging to one of two broad schools of thought. “Formalist” theories claim that humans are born with specific, phylogenetically encoded linguistic knowledge (“Universal Grammar”; Chomsky 1968, 1980) that “bootstraps” children into learning their first language(s), and accounts for the many notable similarities linguists have observed across human languages. “Functionalist” theories hold that linguistic knowledge is learned from information available in children’s ambient environment, and that languages are similar to one another because of common constraints imposed by the properties of humans, the world, and the process of communication (e.g., limits on cognition, biomechanics, information packaging, etc.); if there exist “innate” properties of humans that facilitate this learning, these properties are not specific to human communication (Bates & al. 1998).

The past decade has seen a shift in the focus of research in cognitive and developmental psychology, neuroscience, philosophy, and computer science, away from abstract representations and toward studies of humans interacting with each other and their environment. This collective effort has been identified by some as a movement spanning a range of disciplines, and by others as a new interdisciplinary field of its own (e.g., Ochsner & Lieberman 2001). Findings coming out of this research – including advances in face processing (Kelly et al. 2009), imitation (Meltzoff & Decety 2003), emotion transmission (de Gelder & al. 2006), intention reading (Song & Baillargeon 2008), multimodal perception (Gick & al. 2008), perception-action links (Fogassi & al. 2005), and many other areas – have dramatic implications for the way we think about human language. Yet current linguistic thinking has been largely untouched by the results and methods of this dynamic movement.

The present paper proposes a framework for the study of language with a broad theoretical base in psychology, philosophy and linguistics, firmly rooted in the many recent findings relevant to human interaction. The driving principle behind Interlocution is the notion that humans have evolved – and continue to evolve – in a social and ecological setting in which successful interaction with conspecifics is crucial to survival. While this may seem an obvious point, it has not been a primary consideration in shaping mainstream linguistic theories.

2 Interlocution and mug handles

Humans invented – and continue to invent – language in order to facilitate interaction. As with any human invention, language has inherent design features that fit it to its purpose. In this view, language production is successful only insofar as it is perceivable, and perception only insofar as it connects back to the producer. A powerful analogy to this comes from the literature on the psychology of grasping. Tucker & Ellis’ (1998) seminal study demonstrated that, contrary to the traditional view,

humans do not first perceive manipulable objects in terms of their structural properties and then determine their usefulness; rather, they found that “seen objects automatically potentiate components of the actions they afford.” Thus, the human brain perceives a coffee mug or a hammer first and foremost as a “graspable,” streamlining the perception process in favor of the useful information in the visual scene. The linguistic parallel to this is to say something like: “perceived utterances automatically potentiate components of the communication they afford.” In other words, we do not need to “parse” all of the information in our environment in order to extract what is useful to us. This is an incredibly powerful mechanism with obvious implications for language, and is highly synergistic with the long-standing notion that speech is a “special” kind of signal for humans (Vouloumanos & Werker 2007; O’Callaghan, this volume): speech is not a “natural” stimulus in the human environment, but has evolved to the specific purpose of being produced and perceived by humans. Also, it raises the question of *abstract affordances*: in order to consider the perceptibility of language in terms of how it is useful (i.e., what it “affords”) to humans, it is necessary to characterize what is useful about language.

The “objects” traditionally considered in visual perception have been tangible, physical ones (though we are able to perceive complex, even abstract, properties of these, such as color, orientation, shading, and so on). Likewise, the principal objects of speech perception have typically concerned physical aspects of the world, such as articulatory movements (Fowler 1986) or detectable variations in sound pressure (Diehl & Kluender 1989). In the Interlocution framework, the objects of language perception are those that are truly vital to human interaction, and hence to human survival. Humans are undeniably social creatures, and as such, human survival depends on successful interaction with our interlocutors. This is true even to the extent that humans have evolved the capacity for spoken communication at the expense of a greatly increased likelihood of choking (Holden 2004). Philosophers and social psychologists have recognized that the affordances we perceive in the world (including other humans) cannot be limited to their physical usefulness, but must be extensible to the social (McArthur & Baron 1983) and epistemic (Matthen 2005, this volume). The most essential “objects” one must perceive in human interaction are those at the “highest” level: emotional state, communicative intent, underlying meaning, and so on. Recent literature has shown us that people – even infants and newborns – are surprisingly adept at accessing this kind of information about each other (see references in section 4 below). Likewise, insofar as they are important to successful communication, the same must be true of perceptible units and contrasts at “lower” levels of linguistic analysis (e.g., articulatory gestures, phonemes, words, syntactic structures, etc.).

3 Interlocution, by any means available

In addition to the “special” status of language and the perception of abstract affordances, another keystone of the Interlocution framework is the notion that, as biological creatures evolving in a competitive environment, humans engaging in communication will draw on whatever information is most salient in the available multimodal array of signals. We know that humans can perceive each others’ emotional information from observing either whole bodies or just faces, and can access this information via various combinations of modalities (de Gelder 2006). There is every reason to expect language to draw on this same capacity. For example, our own work has shown that light puffs of air felt on the body can be integrated as speech information in perception, even to the point of overriding auditory perception (Gick & Ikegami 2008).

There is ample evidence that speech is perceived multimodally (McGurk & MacDonald 1976, Gick & al. 2008), even by infants (Aldridge & al. 1999), and that multimodal integration takes place automatically (Soto-Faraco & al. 2004). Presumably the mechanisms of multimodality apply to at least as great an extent in production. The presence of these mechanisms highlights the relative importance of the information, and the relative unimportance of the medium. Further, the most salient information available will normally differ in modality (among other properties) across different individuals (this is most obviously true for impaired speaker/listeners), across producers and perceivers (e.g., direct somatosensory feedback about articulator positions is available to the producer but not the perceiver; Tremblay & al. 2003), across different types of sounds (e.g., the most salient information differs for visible sounds like labials vs. non-visible sounds, for sibilants vs. nasals, etc.), and possibly even in different communicative situations (as in noisy surroundings, D’Ausilio this volume).

The “by any means available” aspect of Interlocution extends deeply into theoretical issues, resulting in a framework in which there is a great deal of room for different approaches to be “right.” For example, the speech literature has been host to a long-standing rift between proponents of auditory-perceptual vs. articulatory objects in speech production and perception. As suggested above, the present framework is consistent with the psychological reality of both of these – indeed, there is excellent evidence for both. Likewise, the framework incorporates concepts typically associated with a realist philosophical position (e.g., affordances) and a representationalist one (e.g., abstract mental structures), while remaining intentionally ambivalent with regard to the deeper philosophical controversy dividing realism and representationalism. The same position is maintained with regard to linguistic formalism vs. functionalism, innateness vs. emergence of linguistic patterns, and so on.

The flip-side of the “by any means available” concept is that not all means are available in all situations. An example of this concerns the atoms of linguistic representation. Phonologists have long disputed whether the appropriate “atoms” of phonological representation ought to be distinctive features (Chomsky & Halle 1968), articulatory gestures (Browman & Goldstein 1992), exemplars (Bybee 2006), etc. There is good evidence that both smaller- and larger-sized “chunks” are needed to produce and perceive language (McQueen & al. 2006). Under Interlocution, the relevance of a particular sized chunk of language depends on the granularity of human cognitive abilities. Thus, at the smallest scale, chunks of language having the size and properties of articulatory gestures are readily manipulable by the human motor system (Kelso & al. 1986), but are unlikely to be stored in long strings in memory. Larger chunks of linguistic information such as words and short phrases, on the other hand, may fit best with the human capacity for episodic memory (Goldinger 1996), which operates over a similar time scale, consistent with an exemplar-based model at roughly the scale of the word.

4 Innateness and Universal Grammar

As described above, formalist theories of linguistics advocate Universal Grammar, an innate complex system of phylogenetically encoded linguistic information, often associated with a language-specific “module” in the mind-brain (Fodor 1983). While the UG proposal is not without its problems (see, e.g., Tomasello 2004), it is unlikely that humans’ specialized skills evolved for social cognition (Herrmann & al. 2007) would not include some innovations specific to language. Under the Interlocution proposal, humans may indeed be born with simple mental structures that constitute an evolved ability to attend and respond to complex properties of human interaction (including language) that occur in their environment – and this set of abilities subsumes what has been traditionally referred to as Universal Grammar. Evidence for the existence of phylogenetically encoded (albeit simple) structured information may be found in at least two perceptual realms, both relating to social and ecological interaction: perception of human faces, voices and emotional expressions, and perception of fear-relevant stimuli (snakes, spiders, heights, etc.). Other inherited specializations must also help to bootstrap children into the language learning process. Some specializations have evolved, at the largest scale, to facilitate interaction in general, others apply only to communication, and still others are (presumably) exclusive to language. Some of these specializations are shared by all social animal species while others are unique to humans.

Classic studies as well as more recent findings on interactive behavior show that newborns (as well as non-human species, Feher & al. 2009) are likely to use a combination of relatively simplistic phylogenetically encoded information, specialized learning mechanisms, and general mechanisms, including: selectively recognizing and attending to human faces (Johnson & al. 1991, Morton & Johnson 1991), imitating facial expressions (Meltzoff & Moore 1977), differentiating others’ emotions via multimodal pathways (Walker-Andrews 2005), preferring speech over other signals (Vouloumanos & Werker 2007), responding selectively to self-, peer- and species-specific cries (Sagi & Hoffman 1976), attending to, imitating, and initiating fine motor movement “dialogues” (Nagy 2006), perceiving speech multimodally (Aldridge & al. 1999), and (within the first few months of life) recognizing and attending to fear-relevant stimuli (e.g., spiders and snakes; Rakison & Derringer 2008). Under the Interlocution framework, the mechanisms that underlie these capabilities in newborns are effectively the same mechanisms by which humans innately attend and respond to structure – including linguistic structure – in their environment. In other words, these abilities, specialized for human interaction and communication, and including some limited structural information, constitute the “bootstrap” infants use

to acquire language quickly and with sparse information.

The proposed mechanisms shared across innate face/emotion recognition, fear response and language imply possible links spanning behavior across these realms. If these abilities are parallel to those needed for language acquisition, we should also expect to see correlations across populations between function in these areas and facility with language. This appears to be the case with both sex-related differences and disordered populations: as to sex-related differences, females (both neonates and adults) tend to be better than males at imitation (Nagy & al. 2007) and more attuned to angry faces (but not to other fear-relevant stimuli; Thunberg & Dimberg 2000), correlating with females' earlier language acquisition. Likewise, for example, autism, a pervasive psychosocial developmental disorder, shows impairment not only to face and emotion perception and fear response, but also language development (Noens & al. 2006, Tager-Flusberg & Caronna 2007) – but not reading (Allott 2001), which separates language comprehension from face-to-face interaction

Regardless of the capacity for phylogenetic encoding of linguistic information in humans – or of the relative simplicity or complexity of that information – the remainder of language must be constructed by each individual, ontogenetically. Thus, whether or not one embraces an innatist view of linguistics, the emergence of linguistic patterns must be accounted for (Archangeli, this volume). The Interlocution framework thus acknowledges a broad range of research programs crossing traditionally functionalist and formalist schools, seeking independent evidence for both the type and complexity of information that may be phylogenetically encodable in humans, and individual variation in linguistic patterns that may be indicative of differently constructed grammars.

5 Abstract structures in Interlocution

Linguists have traditionally considered that the principal function of linguistic structures relates primarily to information processing and storage. Refocusing on the interactive aspects of language, however, highlights another important function for abstract structures in human behavior: predictability. Structure of any kind, and its concomitant predictability, dramatically facilitates the perception process and streamlines the production process. Perhaps even more importantly for human communication, it is only through complex patterns of observable behavior that interlocutors are able to make and test predictions about one another, and thereby generate and test hypotheses about one another's internal states.

As with structure and regularity in any natural system, linguistic structures and regularities enable prediction and confirmation of expectations, allowing most surface details to be ignored or inferred in perception – and a great many to be omitted in production (Hume this volume, Kamide & al. 2003). Thus, for example, in order to perceive a maple tree, one does not need to “parse” all of the available surface information about that tree and then build a complete, fully detailed mental representation of that tree. Rather, someone with knowledge of the general structural properties of trees need only parse a few passing details to adequately perceive the tree. Likewise, the existence of patterned behaviors in language enables humans to attend to – and to produce – only the barest of surface details. This function of enhancing predictive power is a primary purpose of linguistic structure in Interlocution, without which language could not have evolved.

This approach implies that literal parsing and structure building are incomplete and uncomputable aspects of models of perception. Interlocution underscores a deeper problem: While it may seem plausible to parse phonetic information and build, say, a sentence, at some point one must make a mental leap to achieve interpretation – and an even greater leap to achieve insight into one's interlocutor's communicative intent, emotional state, or deeper meaning. Parsing approaches ultimately hearken to what philosophers have termed the Homunculus Argument (Gregory 1987), a mechanism for demonstrating the fallacy of a position by showing that it requires a disembodied perceiver somewhere downstream in the perceptual process to provide interpretation.

An alternative view suggests that humans use other, more robust mechanisms for perceiving each other's inner states. As can be seen from recent literature on human interaction, researchers are beginning to understand some of these mechanisms. In other words, we already know a surprising amount about our interlocutor's intentions (Iacoboni 2005), emotions (de Gelder 2006), and even what he/she is likely to say (Pickering & Garrod 2007). Language enables fine-tuning of our hypotheses about each other,

negotiating between surface observables of language (phonetic detail, usage, etc.) and successful interpretation. A hypothesis-testing component is thus not merely an incidental aspect of Interlocution, but a fundamental one, without which perception could not take place.

An important implication of this aspect of Interlocution is that, because all patterns of interactive behavior find inherent purpose in enabling prediction and hypothesis testing, one need not look further to find functionality. Thus, it is fully predicted by this framework that humans will use their capacity to generate complex patterns to create arbitrary linguistic structures that, at least in some cases, serve no transparent communicative or cognitive purpose other than to facilitate prediction (N. B.: the question of whether the human tendency to create abstract structures and behavior patterns preceded, followed, or developed alongside language, though interesting, is tangential). By way of analogy, the human body is able to move, apparently arbitrarily, to facilitate social interaction (e.g., dancing) or to serve more utilitarian functions (e.g., digging a ditch). In order to understand human behavior, all of these functions must be considered – and as with dancing, in the realm of interactive rather than utilitarian behavior, much of what is observed will appear arbitrary.

6 The neuropsychology of Interlocution

Interlocution implicates brain regions that have previously been of little interest to scholars studying the cognitive neuroscience of language, but which are of obvious relevance for social and communicative interaction. The past decade has seen a dramatic increase in neuropsychological studies of human interaction, implicating several brain regions of interest, e.g.: superior temporal gyrus, the mirror neuron centers (inferior frontal cortex and superior parietal lobe, already considered relevant to language), and the amygdala. The *superior temporal gyrus* is involved in both auditory processing (including language) and social cognition (Bigler & al. 2007). *Mirror neurons*, a subset of premotor neurons which activate in response to the perception of others' specific motor actions (Gallese et al. 1996, Rizzolatti et al. 1996) are also likely an important link in enabling the perception-action link in higher organisms. Mirror neurons have been claimed to be vital to the phylogenetic (Arbib 2005) and ontogenetic (Rizzolatti & Arbib 1998) development of human language as well as other interaction-related functions such as intention-reading (Iacoboni 2005). The *amygdala* is often thought of as an emotion center: part of the brain generally associated with the low-level pick-up of complex, multi-modal cues to emotion and fear-related stimuli (Shaw & al. 2005, Ohman & Mineka 2001). It has been suggested that the amygdala may be linked to language and communication (Schumann 1990, Baas & al. 2004), though these links have not been pursued systematically. De Gelder (2006) identifies the amygdala as a key to understanding multimodal perception of complex, ecologically relevant stimuli: "The amygdala seems to be particularly involved in the assignment of signal relevance... whether these signals consist of sound, sight, smell or touch," suggesting a mechanism whereby "whole-body signals are automatically perceived and understood," reminiscent of our multimodal perception work (Gick & Ikegami 2008).

It is worth noting that autism – with its parallel dysfunctions in social interaction and language function – has been specifically associated with deficits in all three of these regions: the superior temporal gyrus (Bigler & al. 2007), mirror neurons (Williams & al. 2001, Dapretto & al. 2006), and the amygdala (Ashwin & al. 2006). It thus appears likely that there exist neural mechanisms with highly specific functions linking interaction and language – and that these mechanisms are likely to provide a substantial leg-up to language acquisition while also facilitating processing of specific, evolved, ecologically and socially relevant information in the human environment.

7 Conclusion

The Interlocution framework provides an alternative context for linguistic inquiry, linking the formal study of language to other aspects of human interaction. This framework at once establishes strong, largely unexplored links between human capacities for linguistic and non-linguistic interaction, incorporating linguistic structure, speech perception and production, individual and sociolinguistic variation, innateness and emergence, competence and performance, audition, vision and haptics, motor behavior, and so on.

The Interlocution proposal implies a wide range of distinct research programs for language

researchers. Most of these are consistent with approaches currently used in the fields concerned with human interaction and communication. A linguist, for example, might pursue a research program within Interlocution by continuing to analyze the internal and comparative structures of a wide range of languages, and by collecting natural language data to test aspects of the model without needing to become expert in philosophy, speech science, neuroscience, or cognitive, developmental or social psychology. However, as with any broad theoretical framework, Interlocution can expand the horizons and relevance of linguists' work, and can both expand and constrain the kinds of questions linguists ask.

References

- Aldridge, M.A., Braga, E.S., Walton, G.E., and Bower, T.G.R. (1999). The intermodal representation of speech in newborns. *Developmental Science*, 2 (1):42–46.
- Allott, R. *The Great Mosaic Eye: Language and evolution*. Book Guild, Lewes, Sussex.
- Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28(2):105-124.
- Archangeli, D. (this volume). Public and private patterns in language.
- Ashwin, C., Chapman, E., Colle, L., and Baron-Cohen, S. (2006). Impaired recognition of basic negative emotions in autism: A test of the amygdala theory. *Social Neuroscience*, 1:349-363.
- Baas, D., Aleman, A., and Kahn, R.S. (2004). Lateralization of amygdala activation: A systematic review of functional neuroimaging studies. *Brain Res. Rev.*, 45:96-103.
- Bates, E., Elman, J., Johnson, M., Karmiloff-Smith, A., Parisi, D., Plunkett, K. (1998). Innateness and emergentism. In Bechtel, W. and Graham, G (Eds.). *A Companion to Cognitive Science*. OxfordBlackwell, Oxford.
- Bigler, E.D., Mortensen, S., Neeley, E.S., Ozonoff, S., Krasny, L., Johnson, M., Lu, J., Provencal, S.L., McMahon, W., and Lainhart, J.E. (2007). Superior temporal gyrus, language function, and autism. *Dev. Neuropsychol.*, 31:217–238.
- Browman, C. P., & Goldstein, L. (1992). Articulatory Phonology: An Overview. *Phonetica*, 49:155-180.
- Bybee, J. (2006). From usage to grammar: the mind's response to repetition. *Language* 82(4). 711-733.
- Chomsky, N. (1968). *Language and Mind*. Cambridge University Press, Cambridge, MA.
- Chomsky, N. (1980). *Rules and Representations*. Columbia University Press, New York, NY.
- Chomsky, N., and Halle, Morris (1968). *The Sound Pattern of English*. New York: Harper & Row.
- D'Ausilio, A. (this volume). Motor contribution to speech perception.
- de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Rev. Neurosci.* 7: 242-249.
- de Gelder, B., Meeren, H. K., Righart, R., van den Stock, J., van de Riet, W. A. & Tamietto, M. (2006). Beyond the face: exploring rapid influences of context on face processing. *Prog. Brain. Res.*, 155: 37–48.
- Diehl, R. L., & Kluender, K. R. (1989). On the Objects of Speech Perception. *Ecological Psychology*, 1(2), 121-144.
- Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., Iacoboni, M. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience* 9(1): 28-30.
- Fehér, O., Wang, H., Saar, S., Mitra, P. P. & Tchernichovski, O. (2009). *De novo* establishment of wild-type song culture in the zebra finch. *Nature*, 459: 564-568.
- Fodor, Jerry A. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, Mass.: MIT Press.
- Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science*, 308: 662-667.
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, 3-28.
- Gallese, V., Fadiga, L., Fogassi, L. and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119:593-609.
- Gick, B., and Ikegami Y. (2008). The temporal window of audio-tactile integration in speech perception. Paper presented at Acoustics Week in Canada. Vancouver, Canada. October 2008.

- Gick, B., Jóhannsdóttir, K., Gibrael, D., and Muehlbauer, J. (2008). Tactile enhancement of auditory and visual speech perception in untrained perceivers. *Journal of the Acoustical Society of America*, 123(4), EL72-76.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22: 1166-1183.
- Gregory, R. L. (1987). *The Oxford Companion to Mind*. Oxford: Oxford U. Press.
- Herrmann, E., Call, J., Lloreda, M., Hare, B., and Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, 317, 1360-1366.
- Holden, C. (2004). The Origin of Speech. *Science* 303, 1316-1319.
- Hume, E. (this volume). Effects of (un)certainly and expectation on language sound systems.
- Iacoboni, M. (2005). Neural mechanisms of imitation. *Current Opinion in Neurobiology*, 15(6):632-637.
- Johnson, M.H., Dziurawiec, S., Ellis, H.D., Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40:1-19.
- Kamide, Y., Altmann, G. and Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *plus Corrigendum. Journal of Memory and Language* 49: 133-156.
- Kelso, J.A.S., Saltzman, E.L., and Tuller, B. (1986). The dynamical perspective on speech production: data and theory. *Journal of Phonetics*, 14:29-59.
- Kelly, D.J., Liu, S., Lee, K., Quinn, P.C., Pascalis, O., Slater, A.M., and Ge, L. (2009). *Journal of Experimental Child Psychology*, 104(1):105-14.
- Matthen, M. (2005). *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception*. Oxford University Press, Oxford.
- Matthen, M. (this volume). Epistemic affordances: The case of colour.
- McArthur, L.Z., and Baron, R.M. (1983). Toward an ecological theory of social perception. *Psychological Review*, 90:215-238.
- McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264:746-748
- Meltzoff, A.N., and Moore, M.K. (1977). Imitation of Facial and Manual Gestures by Human Neonates. *Science*, 198:75-78.
- Meltzoff, A. N., and Decety, J. (2003). What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society: Biological Sciences*, 358, 491-500.
- McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30(6), 1113-1126.
- Morton, J., and Johnson, M.H. (1991). Conspic and Conlern: a two-process theory of infant face recognition. *Psychological Review*, 98:164-181.
- Nagy, E. (2006). From imitation to conversation: The first dialogues with human neonates. *Infant and Child Development*, 15:223-232.
- Nagy, E., Kompagne, H., Orvos, H., and Pal, A. (2007). Sex-related differences in neonatal imitation. *Infant and Child Development*, 16(3):267-276.
- Noens, I., van Berckelaer-Onnes, I.A., Verpoorten, R., and van Duijn, G. (2006). The ComFor: An instrument for the indication of augmentative communication in people with autism and learning disability. *Journal of Intellectual Disability Research*, 50:621-632.
- O'Callaghan, C. (this volume). Is speech special?
- Ochsner, K.N. and Lieberman, M.D. (2001). The emergence of social cognitive neuroscience. *Am. Psychol.*, 56:717-734.
- Ohman, A., and Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, 108:483-522.
- Pickering, M.J. and Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11:105-110.
- Rakison, D.H. and Derringer, J. (2008). Do infants possess an evolved spider-detection mechanism? *Cognition*, 107:381-393.
- Rizzolatti, G., Fadiga, L., Gallese, V. and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cog. Brain Res.*, 3:131-141.

- Rizzolatti, G., and Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21(5):188-194.
- Sagi, A., and Hoffman, M.L. (1976). Empathic Distress in the Newborn. *Developmental Psychology*, 12(2):175-176.
- Schumann, J. (1990). The role of amygdala as a mediator of acculturation and cognition in second language acquisition. In *Georgetown University round table on languages and linguistics 1990*:169-176. Washington, DC: Georgetown University Press.
- Shaw, P., Brierley, B., and David, A.S. (2005). A critical period for the impact of amygdala damage on the emotional enhancement of memory? *Neurology*, 65:326-328.
- Song, H., Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology*, 44(6):1789-95.
- Soto-Faraco, S., Navarra, J., and Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition*, 92: B13-B23.
- Tager-Flusberg, H., and Caronna, E. (2007). Language disorders: autism and other pervasive developmental disorders. *Pediatr. Clin. North Am.*, 54(3):469-81.
- Thunberg, M., and Dimberg, U. (2000). Gender Differences in Facial Reactions to Fear-Relevant Stimuli. *Journal of Nonverbal Behavior*, 24(1):45-51.
- Tomasello, M. (2004). What kind of evidence could refute the UG hypothesis? *Studies in Language*, 28, 642-44.
- Tremblay, S., Shiller, D.M., Ostry, D.J. (2003). Somatosensory basis of speech production. *Nature* 423:866–869
- Tucker, M. and Ellis, R. (1998). On the Relations Between Seen Objects and Components of Potential Actions. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3):830-846.
- Vouloumanos, A. and Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Developmental Science*, 10(2):159–164.
- Walker-Andrews, A. S. (2005). Perceiving social affordances: The development of emotion understanding. In Homer, B. and Tamis-LaMonda, C. (Eds). *The development of social cognition and communication*:93-116. Erlbaum, Mahwah, NJ.
- Williams, J.H.G., Whiten, A., Suddendorf, T., and Perrett, D.I. (2001). Imitation, mirror neurons and autism. *Neuroscience and Biobehaviour Review*, 25:287–295.

Cognitive and communicative factors in language evolution

Frederick J. Newmeyer

University of British Columbia, Simon Fraser University, and University of Washington

The purpose of this paper is to raise the question of the relative roles of cognitive and communicative factors in the evolution of language. The conclusion is that they have both played important roles, though unbalanced ones. As is evidenced by the basic architecture of grammars, human language originated with the linking of conceptual structures and the vocal output system. In other words, cognitive factors were the first to shape language. But with the passage of time, the needs of communication came to play an evermore important role in grammar. Human language today hence reflects the influence of both types of factors.

1 The difficulty of reconstructing the ancestral human language

The great evolutionary biologist John Maynard Smith once wrote that figuring out how human language evolved is both the most important and the most difficult problem for evolutionary biology (Maynard Smith & Szathmáry 1995). In his view, its importance derives from the truism that language is the most characteristically human of all traits. Hence understanding the origins of language is the key to understanding most other aspects of the history of our species. But why is it a difficult problem? For a number of reasons. First, virtually nothing is preserved over time. There are no archeological digs turning up specimens of the first human language. The fossil record has given us a picture of the evolution of the vocal tract, but grammatical structure is not preserved in geological strata. Second, the comparative method of biology is of no use. This method depends on the existence of homologues to the relevant trait in some related species. But the central aspects of language — syntax and phonology — have no homologues, even for our closest relatives. Third, there is no way to extrapolate backwards to reconstruct an older stage of language. The problem is that the comparative method of historical linguistics allows us to reconstruct ancestral forms spoken at most five to seven thousand years ago. Languages change too fast to go beyond that limit. And most estimates say that *Homo sapiens* is at least 100,000 years old.

Another idea that might seem plausible at first thought would be to take the languages spoken today by pre-industrial peoples — hunter-gatherers and the like — and hypothesize that the first human language was a lot like theirs. After all, it does not seem like an outlandish idea that particular cultural traits would correlate with particular linguistic traits. But in fact, it is impossible to project from the languages of today's pre-industrial peoples to that of the first humans. By the middle of nineteenth century it had become clear that one could carry out the program of reconstructing ancestral sound systems without knowing or caring anything about the culture, society, beliefs, and so on of the people speaking the languages. Even Friedrich Engels, that great polymath of nineteenth century, recognized this point. Engels put forward a socioeconomic explanation for almost everything, but still could write: "It will hardly be possible for anybody, without being ridiculous, to provide an economic explanation for the Germanic sound changes" (cited in Murra, Hankin, & Holling 1951: 60). In other words, linguists had discovered that grammatical structure is not intrinsically linked to sociocultural aspects of speakers. If one slogan captures the essence of twentieth century linguistics, it is an impassioned statement from Edward Sapir, the greatest North American linguist of the early half of the century: "When it comes to linguistic form, Plato walks with the Macedonian swineherd, and Confucius with the headhunting savage of Assam" (Sapir 1921: 219). The driving sentiment of modern linguistics is that all languages are equal in four crucial respects. First, there is no such thing (grammatically speaking) as a primitive language or an advanced language. Second, grammars are cut from the same mould. That is, grammars of all languages are composed of the same sorts of units — phonemes, morphemes, and so on. Third, it therefore follows that all grammars can be analyzed by means of the same theoretical tools. And fourth, for any given language, there is no correlation between aspects of that language's grammar and properties of the users

of that grammar. Put simply, all of the 6000 or so languages of the world are pretty similar. Given the logically possible ways that structured communication systems could differ from each other, there are not really that many differences. The same grammatical devices turn up in language after language after language. When linguists go out in the field to describe a new language, they are sometimes surprised to discover some novel grammatical feature. But more often than not, what they find are variations on familiar themes.

Let us try a different strategy for untangling the origins of language. This strategy begins with the question of what language is used for. The standard immediate answer is ‘communication’, and of course that is correct. We interact with each other and language is the primary medium for doing so. But language has another crucial function too, namely that of allowing us to mentally represent the world: a cognitive role, if you will. The origins of language debate is one more arena between the two major opposing camps in the field of linguistics. Functionalists have tended to emphasize the ‘design’ of language for communication, whereas formalists have tended to emphasize its ‘design’ for cognition (for discussion, see Newmeyer 1998). My principal argument is that both functionalists and formalists are (partly) right. Language developed first to support cognition and only later as a medium of communication.

2 Communicative influences on grammar

There are many ways that language supports communicative needs or, otherwise put, that it seems designed for efficient use. Perhaps the most important is that grammars are designed so that hearers can figure out the meaning of what they hear as rapidly as possible. There are two ways that grammars accomplish that. The first is by lining up the elements of the sentence in order of complexity. Take a language like English, where objects follow their heads:

- (1)
 - a. Mary refused the offer. (verb - object)
 - b. Mary lived in Hull. (preposition - object)
 - c. Mary is fond of tennis. (adjective - object)
 - d. Mary’s refusal of the offer. (noun - object)

In each case a ‘shorter’ head precedes a ‘longer’ object. But the short-before-long tendency in English grammar involves far more than the head-object relation. All grammatical categories after the verb line themselves up in short-to-long order, as the following sentence illustrates:

- (2) $_{VP}$ [convince - my students - of the fact - that linguistics is interesting]

Also notice that single adjectives and participles can appear before the noun:

- (3)
 - a. a silly proposal
 - b. the ticking clock

But if these adjectives and participles themselves have modifying material following them, that material has to appear after the noun:

- (4)
 - a. *a sillier than any I’ve ever seen proposal
 - b. a proposal sillier than any I’ve ever seen
- (5)
 - a. *the ticking away the hours clock
 - b. the clock ticking away the hours

Also, where speakers have a choice in English, they will typically choose short-before-long. Both (6a) and (6b) are possible sentences of English, but in actual communication speakers are far more likely to say the latter than the former:

- (6)
 - a. [That the train will leave on time] is unlikely.

- b. It is unlikely [that the train will leave on time].

Studies indicate that grammars are easier to use when the parts of sentences are lined up in order of increasing or decreasing complexity, a fact that seems to indicate that grammars are designed for the people who *use* language (see Hawkins 2004 for discussion).

There is another way that grammars seem designed for language users. In general, what we find is an iconic relationship between form and meaning (Haiman 1985). For our purposes, that means that the form, length, complexity or interrelationship of elements in a linguistic representation reflects the form, length, complexity or interrelationship of elements in the concept that that representation encodes. Consider two examples. The first involves causation in English, which can be expressed in two ways:

- (7) a. The lexical causative: *kill, persuade, melt, ...*
b. The periphrastic causative: *cause to die, cause to believe, cause to melt, ...*

Lexical causatives tend to convey a more direct causation than periphrastic causatives. Note that (8a) implies that Bill's death was a more direct result of the stabbing than does (8b):

- (8) a. John killed Bill by stabbing him in the heart.
b. John caused Bill to die by stabbing him in the heart.

Now consider (9a-b):

- (9) a. John caused Bill to die by buying him a ticket on a flight that ended up being subject to a terrorist attack. (*indirect*)
b. ???John killed Bill by buying him a ticket on a flight that ended up being subject to a terrorist attack. (*direct*)

The strangeness of (9b) derives from the fact that Bill's lack of knowledge about the attack while buying the ticket is incompatible with the direct causation implied by the word *kill*. So we can see that where cause and result are formally separated, conceptual distance is greater than when they are not.

Another example can be drawn from the concept of possession. There are two important types of possession in human language.

- (10) a. Alienable possession: *John's book*
b. Inalienable possession: *John's heart*

In English they have the same structure. But in a majority of languages it is more complicated to say 'John's book' than to say 'John's heart'. And in no language in the world is it more complicated to say 'John's heart' than to say 'John's book'.

Also, there are many properties of grammars that seem irrelevant to cognition, yet are good for communication. First and most importantly, language has phonology. Language is — typically — spoken. And many phonological changes over time seem designed to make things easier for the speaker or the hearer or both. Second, language has morphology. The use of prefixes and suffixes helps to shorten the time it takes to say frequently used ideas. Consider (11a-b):

- (11) a. That is a **non**-issue
b. Bill is work**ing**

We could convey the same ideas without affixes, but it would take longer. Third, languages ease comprehension by marking clause boundaries with complementizers:

- (12) I think *that* it is time to leave.

Fourth, complex semantic notions tend to be conflated into the two grammatical relations ‘Subject’ and ‘Direct Object’:

- (13) a. Mary threw the ball. [‘Mary’ is the Agent of the action]
 b. Mary saw the play. [‘Mary’ is the Experiencer of an event]
 c. Mary received a letter. [‘Mary’ is the goal/recipient of transfer]
 d. Mary went from Chicago to Detroit. [‘Mary’ is an object undergoing transfer of position]

Even though its semantic role differs in each sentence, *Mary* acts as the grammatical subject. If language were designed just for cognition, we would expect the different meanings to be kept distinct grammatically. And fifth, grammatical elements are often displaced from the position where they are understood:

- (14) Who did you talk to ____ ?

Putting the *who* at the beginning of the sentence focuses right away that it is a request for information, even though the consequence is the breaking up of a semantic unit.

3 Grammars as reflections of cognition

Now let us consider the opposite situation. In fact that there are many ways that grammars seem poorly designed for efficient communication, but well designed from the standpoint of cognition. First, every language on earth allows for the possibility of recursion, that is, structures embedded inside of structures inside of structures, ad infinitum. For example, in principle, there is no limit to the number of times that another subordinate clause can be added in sentences like the following:

- (15) Mary thought that John said that Sue insisted that Paul believed that ...

Is recursion necessary for communication? Apparently, it is not. We virtually never have any reason to utter complex sentences like (15). And the desired message conveyed by simpler sentences with recursion like (16a) can easily be communicated by a sentence like (16b), employing juxtaposition of two clauses:

- (16) a. Mary thought that John would leave.
 b. Here is what Mary thought. John was going to leave.

Actually, everyday communication makes use of surprisingly little recursion. Givón (1979) and many others have suggested that the use of subordinate clauses increases dramatically with literacy. But the fact that every language allows the possibility of recursion suggests that it is a genuine design feature of language, there from the beginning. Why would this be the case? The obvious answer is that human thought has recursive properties, even if the manifestation of the expression of that thought in communication does not necessarily draw on those properties. As noted by Pinker and Jackendoff:

Indeed, the only reason language *needs* to be recursive is because its purpose is to express recursive *thoughts*. If there weren’t any recursive thoughts, the means of expression wouldn’t need recursion either. (Pinker & Jackendoff 2005: 30; emphasis in original)

Second, consider the rampant structural ambiguity that grammars allow. I do not think that it is controversial that the more ambiguity a communication system tolerates, the less efficient it is. By that criterion, human grammars are horribly designed for communication. Virtually any sentence imaginable is loaded with potential ambiguity. For example, sentence (17) was calculated by Martin, Church, and Patel (1987) to have 455 parses:

- (17) List the sales of products produced in 1973 with the products produced in 1972.

Of course, we have found ways to deal with this problem in actual language use. For that reason, humans have developed complex systems of inference and implicature, conveyed meanings, and so on. Hence, in conversation, real ambiguity is normally a minor problem. But our concern here is the design of grammars and whether they are well shaped for communicative purposes. Based on the ambiguity that they permit, the conclusion has to be that they are not well shaped. If we focus on cognitive representations instead of on communication, however, structural ambiguity is a much less serious problem. The reason is that many (communicatively) ambiguous sentences are disambiguated by their structures. Consider for example the following ambiguous phrase:

(18) The old men and women (applied for senior citizen discounts).

In actual speech, the scope of 'old' could be simply 'men' or be 'men and women' understood collectively. But this phrase is disambiguated in terms of its structural (i. e. cognitive) representation. The former representation is as in (19a), the latter as in (19b):

- (19) a. [the old men] and [women]
b. the old [men and women]

From the point of view of language use, the possibility of dual (i.e. ambiguous) representations for the same sequence of words is not communicatively desirable. But since the different meanings are represented differently from the cognitive standpoint, we must conclude that in this respect grammars seem well adapted to cognition.

Third, consider the possibility of communicatively useless sentences. A central fact about language that it allows us say anything that we can conceptualize, regardless of whether we would actually have any need, desire, or likelihood to convey the information conceptualized. No more effort is required to say an obviously false sentence (20a) or an obviously true one (20b), than one that might be genuinely communicatively relevant (20c):

- (20) a. Tomorrow will not follow today.
b. Tomorrow will follow today.
c. Tomorrow will be rainier than today.

Along the same lines, we have no more trouble uttering pure nonsense sentences like the celebrated (21a) than grammatically parallel ones that contain an easily accessible semantic content (21b):

- (21) a. Colorless green ideas sleep furiously.
b. Revolutionary new ideas appear infrequently.

Any speaker of English knows that (21a) is absurd and knows why it is absurd. That is, we assign to it a conceptual representation which registers the incompatibility of being simultaneously 'colorless' and 'green', the impossibility of ideas have (literal) color, the incongruity of sleeping in a furious manner, and so on. But none of that prevents our being able to utter (21a) with the same facility as (21b). In other words, communicatively useless sentences provide another example of how language is 'overdesigned' for communication.

Sometimes, of course, there are obvious functional reasons for why all languages are so similar. That is, there are explanations based on pure usefulness. It does not require a sophisticated theory to explain why every language can express the concepts 'sun' and 'moon', 'mother' and 'father', 'heart' and 'lungs', and so on. And everybody needs to express concepts like negation, to ask questions, to give commands, to distinguish in some way between things that happened in the past and might happen in the future. It is hardly a surprise, then, that we find grammatical means in every language to express these concepts. Likewise, we are not surprised to find that no language is completely missing vowel sounds and that no language has 100 distinct sounds. Languages like those would be impossible to pronounce, to understand, or both.

But what makes it easy to show that language is not just based on communication are all of the things that no language in the world does, and where there is no explanation for the absence of the property that is plausibly based in communicative utility. For example, there is no language in the world where words are entirely made up of a sequence of sounds, each of which corresponds to some aspect of the meaning of the word. So there is no language where names of physical objects, say, all start with [k], names of animals all have an [o], and names of mammals all have a [b], and so on, so that the word for 'hedgehog' would start [kob]. What is interesting about this impossibility is that for hundreds of years many made up languages have had just that property. In other words, there is nothing off the cuff implausible about languages like that. They certainly might be useful, but they just do not exist. Second, there is no language in the world where negatives, questions, or commands are systematically formed by dropping the first sound of the verb. That is, there is no language where "Finish your supper!" would be said "Inish your supper!", where "Call me tomorrow" would be said "All me tomorrow," and so on for every command in the language. Third, many languages form questions by doing nothing but changing the intonation contour, that is, the pitch of the sentence. So in English, we can question the idea of your leaving by saying "You're leaving???", with rising pitch on the verb. But in no language are *negatives* formed by changing the intonation contour. Fourth, there is no language where the syntax can 'count', or at least no language where it can count past two. For example, we never find languages where, say, one negates an affirmative by sticking in a word like 'not' as the third word of the sentence. In some languages, like Spanish, the negative goes pretty regularly before the verb. In other languages, like German, the negative goes pretty regularly after the verb and the object. In other languages, like English and French, negation is a lot more complex. But we never find negative words being stuck in precisely after the third word, or the fourth, or the fifth. This circumstance is quite odd. People can count easily enough to three or four or five, so why not their grammars? And finally, there are some curious gaps in vocabulary across languages as well. Most languages have words that translate as *all*, *some*, *both*, *none*, and so on. But no language known has a word that translates as *not all* (Horn 1989). So in no language can one say something like *Nall the children* are here, meaning 'Not all the children are here'. Languages in fact seem not to have words for logical complements of any sort, that is, words that convey the idea of 'everything except for something'. So one could imagine that a language might have a word, *guhthree* that would mean 'all but three', *guhfour* that would mean 'all but four', *guhchildren* meaning everyone except the children, and so on. But languages do not do that.

It is the fact that all the languages of the world do certain things, even though there is no communicative reason why they should, and the fact that no languages in the world do certain things, even though there is no communicative reason why they should not, that form the meat of grammatical theory. This circumstance has led linguists to suggest that the human brain is hardwired for particular grammatical properties (Chomsky 1986) and has resulted in the biological evolution of language being a major topic of research.

So if there is some respect in which all of the languages of the world behave in the same way, that could be in part a consequence of the fact that all humans share the same body plan, as well as sharing certain needs and experiences. But it probably is also a consequence of an innate Universal Grammar that shapes the grammars of all languages, and which is only indirectly related to communicative usefulness.

4 Cognitive influences preceded communicative ones

Now, where are we? We have seen several important design features of language that pertain little — if at all — to communication. And at the same time they seem part and parcel of cognition. Thought is recursive, so it is not surprising that grammars are. Ambiguous sentences that might cause difficulty for communication are disambiguated mentally. We can say what we can conceptualize, regardless of whether we might ever need to say these things in real life. And all of these features occur in every language. That suggests that they were there from the start, that is, at the dawn of human language itself. But now consider those aspects of language that seem designed to better aid communication. They tend to be much less central to the basic design plan of language. For one thing, languages tend to manifest them in wildly different degrees. It is certainly easier for speakers and hearers if verb phrase structure goes consistently from short to long (as in English) or from long to short (as in Japanese). But there are a

significant number of ‘inconsistent’ languages where one observes a mixture. So in German, the noun comes *before* its object, but the verb almost always comes *after* the object:

- (22) a. den Apfel essen ‘the apple eat’
b. das Essen des Apfels ‘the eating of the apple’

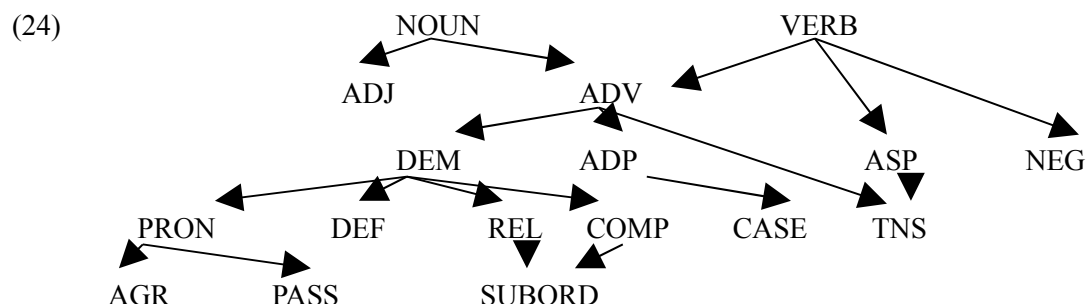
To take another example, morphological particles like prefixes and suffixes are certainly useful to rapid communication. But quite a few languages — Vietnamese is one — have no affixes at all. Even phonology is not universal, since only spoken languages manifest it (one does refer to the ‘phonology’ of signed languages, but the principles governing it are quite different from those governing spoken languages).

Another fact about the communication-enhancing aspects of language is their historicity. In many cases one can observe them developing over time and reconstruct their ancestors in categories more directly reflecting cognition. Consider a communicative aspect of language par excellence, namely, discourse markers, that is, expressions like:

- (23) *then, I mean, y’know, like, indeed, actually, in fact, well, ...*

Discourse markers are essential to the makings of a coherent discourse. Yet no other phenomenon could be as historical in the literal sense of the term. That is, they invariably arise from something else, typically out of more directly conceptual meanings (Traugott & Dasher 2002). For example, *then* comes from the adverb, *y’know* from a full sentence, and so on. This circumstance seems at first blush quite curious. Why would they be central to communication, but derivative historically? But if vocal communication itself is derivative, it all makes sense. Nouns and verbs trace back to nouns and verbs, because they were there from the start. The derivative nature of discourse markers points to a time when we had structured conceptual representations, but they had not yet been co-opted for communication.

It is not just discourse markers that seem derivative, evolutionarily speaking. By and large, elements that convey grammatical meaning and those that supply the clause with aspectual modalities, nuances of quantification and reference, and so on invariably derive from something else, and ultimately from nouns and verbs. Heine and Kuteva (2002) have illustrated the most common pathways by which such elements arise as in (24):



ADP=adposition; ADJ=adjective; ADV=adverb; AGR=agreement; ASP=aspect; COMP=complementizer; DEF=definite marker; DEM=demonstrative; NEG=negation; PASS=passive; PRON=pronoun; REL=relative clause marker; SUBORD=subordination marker; TNS=tense

Heine and Kuteva draw the reasonable conclusion that in the earliest form of human language there were only two types of linguistic entities, one denoting thing-like time-stable entities (nouns, in other words), the other denoting actions, activities, and events (verbs, in other words). While they themselves do not go on to suggest these facts point to a pre-communicative stage for language, the conclusion seems like a reasonable one.

There is another reason to posit that cognition left its mark on language before communication. We have learned that the conceptual abilities of the higher apes are surprisingly sophisticated. Each

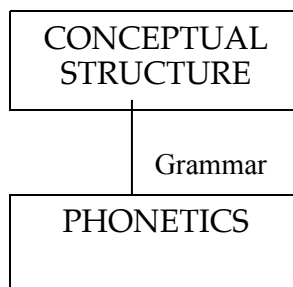
passing year leads to new discoveries about their capacity for problem solving, social interaction, and so on. However, their communicative abilities are remarkably primitive. There is very little calling on their conceptual structures in communicative settings. These facts, taken together, suggest a three-stage process in language evolution. First, there existed a level of conceptual structure:

(25)



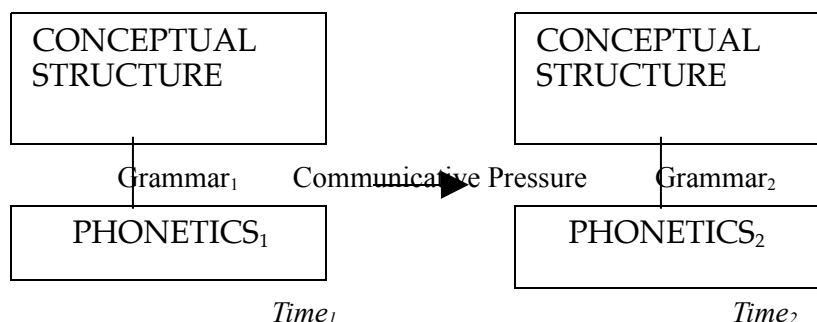
Secondly, the level became linked to the vocal output channel, creating for the first time a grammar that is independent of the combinatorial possibilities of conceptual structure per se and making possible the conveying of thought — in other words, communication. Example (26) illustrates:

(26)



However, once grammars started to be drawn upon for real-time purposes, the constraints of real-time use began to affect their properties. The development of phonological systems is perhaps the most obvious example. Grammars over time began to be shaped to facilitate processing in various ways. This is most evident in their shaping to allow the more rapid expression of frequently-used meaningful elements than of those less frequently used ones. As a result, it is auxiliaries and negative elements that tend to contract in English, rather than full lexical nouns and verbs. Many languages developed affixes for the most commonly-used concepts, such as negation, causation, comparison, and so on, but rarely for more complex infrequent concepts. The dozens of different semantic roles expressible became packaged into a small number of grammatical relations such as ‘Subject’ and ‘Direct Object’. Displacements, such as are illustrated in (14), arose to focus or to downplay the importance of constituents for the discourse, driven by communicative considerations. (27) represents the historical process I am describing:

(27)



To conclude, reconstructing in detail the properties of the first human language seems pretty hopeless, at least for the foreseeable future. Still, there are conjectures about our ancestors’ language that seem pretty safe to make. Human language was jump-started by the linking of conceptual structures and the vocal output system. In other words, cognitive factors were the first to shape grammars. But with the passage of time, the needs of communication came to play an evermore important role in grammar. Human language today reflects the influence of both types of factors.

References

- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Praeger, New York.
- Givón, T. (1979). *On understanding grammar*. Academic Press, New York.
- Haiman, J., ed. (1985). *Iconicity in syntax. Typological Studies in Language, vol. 6*. John Benjamins, Amsterdam.
- Hawkins, J.A. (2004). *Efficiency and complexity in grammars*. Oxford University Press, Oxford.
- Heine, B., & Kuteva, T. (2002). *On the evolution of grammatical forms. The transition to language*, ed. by Wray, A. 376-97. Oxford University Press, Oxford.
- Horn, L.R. (1989). *A natural history of negation*. University of Chicago Press, Chicago.
- Martin, W.A., Church, K.W., & Patel, R.S. (1987). Preliminary analysis of the breadth-first parsing algorithm: Theoretical and experimental results. *Natural language parsing systems*, ed. by Bolc, L. 267-328. Springer Verlag, Berlin.
- Maynard Smith, J., & Szathmáry, E. (1995). *The major transitions in evolution*. W. H. Freeman, Oxford.
- Murra, J. V., Hankin, R.M., & Holling, F., eds. (1951). *The Soviet linguistic controversy*. King's Crown Press, Oxford.
- Newmeyer, F.J. (1998). *Language form and language function*. MIT Press, Cambridge, MA.
- Pinker, S. & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, 95:201-36.
- Sapir, Edward. 1921. *Language*. Harcourt, Brace, and World, New York.
- Traugott, E.C., & Dasher, R.B. (2002). *Regularity in semantic change*. Cambridge University Press, Cambridge.

Epistemic affordances: the case of colour*

Mohan Matthen
University of Toronto

1. What is an Epistemic Affordance?

In this paper, I attempt to solve a problem concerning our knowledge of colour by reorienting J. J. Gibson's (1979) concept of affordance (chapter 8).

The core of Gibson's notion is, as I take it, that animals sense objective real-world objects and properties, but sense them in subjective terms – terms directly relevant to their own possibilities for “behaviour” (as Gibson puts it). Air and water are real-world objects, and we sense them.¹ When we sense (or seem to sense) air and water, our senses are objectively right or wrong. For example, when we experience a mirage, we seem to sense water, but the sense-experience is wrong – *objectively* wrong. In the same vein, if a body of water looks like ink (because of odd illumination, say), then our experience of it is mistaken. In other words, *water* – real-world objective water – is a reference point against which experience-as-of-water has to measure up for accuracy and verisimilitude.

The argument just given establishes that the senses inform us of things, and of states-of-affairs, that are objectively real. Gibson wants to emphasize, however, that when we sense such things, we do not sense them in terms that fit smoothly into the descriptive vocabulary of an observer-neutral science such as physics. Rather, we sense them in terms of the possibilities for behaviour that they “afford” *us*. (They may not afford a differently constituted organism the same possibilities.) Air, for example, is a medium that affords us breathing and unimpeded passage, provided that our feet are planted on solid earth. (For a bird, the proviso is unnecessary.) Water is a substance that shares a boundary with air or solid, which we can pour and drink, but which is not a medium for ambulation. (For a fish it is more like air is for a bird, and does not always have a boundary.) A flat, horizontal solid surface affords us support for ambulation; a vertical solid is a barrier that has to be circumambulated. Such descriptions, which Gibson took to be the products of the sensory process, show that sensation does not offer us a view of “a value-free physical object to which meaning is somehow added,” but rather “a value-rich ecological object.” We sense objects as apt for the behaviours mentioned.

This is the view of affordance that I want to reorient here. I shall argue that it is extremely useful in understanding why we possess some little noticed certainties in sensory knowledge that I shall outline in a moment. At the same time, I find its emphasis on *behaviour* unduly limiting. By ‘behaviour’, Gibson has in mind body-world physical interactions – the kinds of bodily movement that allow you to move around in, or act upon, your surroundings. The kinds of activity he mentions are: *breathing, walking, eating, throwing, grasping*, and so on. But he does not mention *deciding, remembering, enjoying, thinking*, and other mental activities. When one sees water, one does not merely sense that it is a no-walk area – one also remembers where it is, enjoys its presence, and so on. Why then does the sight of water not equally afford such mental actions?

As far as I am aware, Gibson does not explicitly argue in favour of his narrow construal of affordance, but it is not uncommon for psychologists and others to assume that the sensory guidance and control of bodily motion have some kind of evolutionary priority over the role that the senses play in mental life. Recall Patricia Churchland's witty line: “Boiled down to its essentials, a nervous system enables the organism to succeed in the four F's: feeding, fleeing, fighting, and reproducing.” It is hard to resist the idea that something of this sort lies behind the Gibsonian insistence on behavioural affordances.

* Parts of this paper are taken from my “The Sensory Representation of Colour”, in Jonathan Cohen and Mohan Matthen (eds) *Colour Science and Colour Ontology* (Cambridge MA: MIT Press, forthcoming 2010).

¹ As the title, *The Senses Considered as Perceptual Systems* [1966] intimates, Gibson does not make the usual distinction between sensation and perception: for him, sensing is not merely an internal event, but an intimation of external happenings.

This attitude overlooks the ancient evolutionary origins of epistemic activities. Sensation affords us the capacity to create, build up, and modify a *retained internal record* of our external surroundings. Sensation *conditions* us to external associations; it *primes*, *sensitizes* and *habituates* us to types of events. In these and other ways, our behavioural patterns are changed by what we have sensed. These changes constitute records of what we have sensed in the past. The creation, modification, and deletion of these internal records are just as much functions of perception as bodily control. But they are not themselves “behavioural”, in the bodily sense. Record-keeping – understood broadly enough to include the modification of behaviour – is a characteristic of extremely primitive organisms.

It is against this background that I introduce the notion of an *epistemic affordance*. The idea is that sensation affords us the possibility of building *knowledge*. In the most primitive way, it modifies an organism’s reactions to environmental events in response to information about the nature of these events. Take habituation: in response to a stimulus occurring repeatedly, an organism begins to take it as a regular part of its environment, and thus pays it less attention. Or classical conditioning: in response to two stimuli being regularly associated, an organism reacts to one when it senses the other. My point, extending Gibson, is that sensation affords us the possibility of interpreting previously encountered stimuli in these ways.

More sophisticated organisms, of course, have more complex and indirect knowledge-gathering operations – discursive reasoning, induction, theory-construction, etc. But the relatively restricted extent of these sophisticated operations, which philosophers understandably tag as paradigms of knowledge-gathering, should not blind us to the evolutionary priority of the more primitive operations I have mentioned, nor to the evolutionary continuity of the more sophisticated operations with the more primitive. Theory construction is a paradigm of rationality; conditioning is sub-rational –nonetheless both are forms of record-keeping, or, to put it more grandiosely, of knowledge-construction. The senses afford us both.

2. Sensory Object-Knowledge

I take as my starting point some certainties contained in sensory knowledge. I shall be considering colour as an example, though the points I make can be repeated for other sensory qualities. I shall try to show how Gibson’s notion of an affordance, expanded to include epistemic affordances, helps us understand these certainties.

Consider a statistically normal human colour perceiver, Trich (so-called for her trichromacy) viewing a piece of fruit in reasonably good light. Suppose that to Trich this fruit looks orange. Trich can take her visual state *O* as strong support for the proposition:

OK (1) That [fruit] is orange.

Now, this kind of visual state is rarely the whole story about sensory knowledge. For as Gibson says:

The higher animals have evolved both mobile extremities and adjustable sense organs. Hence they can modify the stimulus input in two ways: by moving the organs of the body that are called ‘motor’ and by moving the organs of the body that are called ‘sensory’. . . . [S]ome movements accomplish behavior in the usual meaning of the term and other movements accomplish the pickup of sensory information. The former will be called performatory or executive, the latter exploratory or investigative. . . . Animals and men can select or enhance the stimuli they receive from the world or even exclude certain kinds, by orienting and adjusting their sense-organs (1966, 32).

If Trich is *examining* the fruit, she will not stop at *O*. She will take the fruit to the window, where the sunlight is stronger; she will put her glasses on; she will try to eliminate any sources of non-standard illumination; she will turn the fruit over in her hands or bring it closer to her eyes. She will, in short, engage in *sensory exploration*. At the end of this process, she would be in a state of *empirical* certainty

about (1). By *empirical certainty*, I mean this: Trich has no *isolated* reason for doubting (1) – no reason that would not simultaneously lead her to doubt a whole raft of unconnected propositions.

Empirical certainty does *not* entitle her to dismiss *sceptical* doubt. She may be dreaming, deluded, or deceived by an evil demon; she may be a brain in a vat; her visual pathways may be subject to a strange form of transcranial interference. This kind of doubt is not, however, *isolated*. Since sceptical doubt attacks background conditions for empirical knowledge, it attacks not only (1), but equally many other propositions. The possibilities just mentioned would lead her in turn to doubt whether the lemon next to the orange was really yellow, whether it was really her hand in which the fruit resided, whether it was really daylight, whether she was really awake, and so on. Sceptical doubt spreads by undermining the foundations of empirical knowledge. Empirical certainty is subject to the security of these background conditions.

My point is that given careful visual inspection, Trich may still doubt (1), but only in an expansive way. More generally: after a perceiver has undertaken the full range of exploratory activities open to her, she has no reason to doubt what her senses tell her, other than reasons that apply to unrelated and independent facts. This is part of the nature of sensory knowledge. It makes no sense to say that something has a sensory feature (e.g. orange), but does not appear as if it has that feature in *any* circumstance of viewing that the observer can get herself into.

Empirical certainty is a peculiar outcome of sensory exploration. We are more certain of things we sense than of the theories that we mentally construct. This is because the meaning of sensory knowledge is closely tied to activities of sensory exploration. The *meaning* of saying that something is blue is closely tied to the activities and experiences by which I ascertain the colour of things. On the other hand, the external world is ultimately a check or reference point for determining the truth of sensory knowledge. This is why one cannot be *completely* certain of sensory knowledge – one cannot dismiss sceptical doubt. These are the two aspects of sensory knowledge that I'll be discussing in this paper. Let's mark them by saying that sensory knowledge has *limited immanence* – enough immanence that it has a higher degree of certainty than theoretical knowledge, not enough to give it Cartesian certainty.

3. Sensory Feature Knowledge

Let us move on now to a second kind of certainty. In light of her experience, *O*, Trich is also entitled to certainty regarding certain propositions about the *colour* presented therein. To wit:

CK (2) The colour presented in *O* is yellowish.

CK (3) The colour presented in *O* is less reddish than that presented in another visual state *C* (which happens to be occasioned by a cherry).

CK (4) The colour presented in *O* is less yellowish than that presented in another visual state *L* (which happens to be occasioned by a lemon).

CK (5) The colour presented in *O* is more similar to that presented in *L* than it is to that presented in *C*.

These propositions concern the *colours* that Trich is acquainted with, not the objects that possess them.

Now, Trich's certainty with regard to CK propositions is actually higher than empirical certainty. For, it seems that to know the above sort of fact about the colours, one needs only to experience them.² For Trich to know that the colour of the thing she seems to see is yellowish, as asserted in (2), she needs nothing more than her experience of that colour. She does not need to know that the fruit actually is the colour it seems to be; she does not even need to know that the fruit exists. Even if Trich's experience, *O*, is a dream or a hallucination, it is an experience of a certain colour, and this is enough for her to be sure

² This point is the converse of Descartes' claim that one can clearly and distinctly conceive of a chiliagon, but not know it by perceiving it.

that the colour is yellowish in hue. Hallucinatory colour-experiences are no less probative with regard to the colour-knowledge of the sort contained in (2) – (5) than normal veridical experiences. For this reason, Trich is entitled to dismiss even sceptical doubt with respect to the CK propositions. She is entitled to *Cartesian certainty* with regard to them. Cartesian certainty is proof even against sceptical doubt.

We can sum up the results of this section by saying that sensory feature knowledge is *completely immanent*.

4. Experience-Dependent Accounts of Colour

Some theories of colour link it essentially to sensation. These theories are very different from a Gibsonian account, which emphasizes not just the perceiver-relativity of sensation, but also its objective significance. How would an experience-based theory account for the evidentiary status of colour vision as just outlined?

Consider, for starters, a theory that identifies colour with colour-sensation. Since such a theory is extremely implausible, I'll be brief. We are aware of the character of our own experiences. There can be no doubt about their character. If the CK-propositions are about the character of experience, then they admit of no doubt – experience is completely immanent. Thus, the experience-dependent account squares with the Cartesian certainty of these propositions. But these theories stumble on object knowledge. If Trich's experience of the colour of an object is *sufficient* for the object's having a certain colour, then she should be as certain of object-colour as she is of the character of colour itself. But she is not: as we have seen she has only empirical certainty of object-colour, though she has Cartesian certainty concerning what she knows about colour itself. Plausibly, she lacks the ultimate degree of certainty about object-colour exactly because it is based on something real. To put in another way, the very utility of sensory exploration of colour suggests that there is something out there to explore. This is what a purely subjective account of colour cannot accommodate. It cannot accommodate the *limited* immanence of knowledge gained through sensory experience.

Let's consider a more plausible example of a subjectivist theory. According to Dispositionalism, colour is a disposition in things to evoke in "normal" perceivers a certain type of sensation in "normal" circumstances. According to this theory, a fruit is orange in virtue of the fact that it possesses a disposition to create a certain type of colour sensation, of which Trich's visual state *O* is an example. Dispositionalists suppose that the Cartesian certainty of Trich's colour-knowledge arises from the fact that she has incorrigible access to her own sensations, which have the properties in question – reddishness, yellowishness, etc. Facts concerning the compositionality of colour as in (2) – (4) and similarity relations as in (5) ride on some quality of the sensation, according to the Dispositionalist, and are thus knowable with certainty.

Dispositionalists have a partial account of Trich's merely empirical certainty with regard to object-knowledge. When she first looks at the fruit, Dispositionalists say, she is certain only of the sensation that the fruit evokes in herself in her particular circumstances of viewing. This is *some* evidence that it possesses a disposition to evoke this sensation in normal circumstances, but not conclusive evidence. By examining it closely in a variety of normal viewing conditions – by bringing it to the window, turning it over, and putting on her glasses – she expands the range of circumstances in which she has viewed the fruit. Thus, she increases her certainty about its dispositions with respect to her own sensations in normal circumstances of viewing. However, such certainty is subject to sceptical doubt – this kind of scrutiny does not rule out the possibility that she is systematically deluded about her circumstances.

At first sight, then, Dispositionalists can account for the distinctive kinds of certainty possessed by the OK and CK propositions. There is, however, an obstacle on the dialectical path to object knowledge – Trich's empirical certainty regarding the colour of something she sees. According to Dispositionalists, the colour of the fruit is its disposition to create a certain kind of sensation not only in Trich, but in other normal perceivers as well. Looking at the fruit in all kinds of different circumstances certainly helps Trich know its propensities with regard to her own colour experiences – but it does not help her know how it is disposed with regard to other perceivers. Examining the fruit closely could not help her much with this; this procedure only gives her broader knowledge of the fruit's dispositions to

evoke sensations in herself. (In fact, as we shall see, it is most likely *false* that things elicit exactly the same sensations in different perceivers.) To draw any conclusion about what sensations it creates in other perceivers requires other arguments – including, perhaps, an argument by analogy. No such argument is compelling enough to provide Trich with anything like empirical certainty. Knowledge of *other* people's experiences is not immanent.

5. Primary Quality Accounts of Colour

One can be objectively wrong about what colour something is. This suggests that there is a real-world check on colour experience. So it seems that colour has a real-world presence. This motivates primary quality theories of colour. According to these theories, colour experience is no part of the essence of colour. An example of a position of this kind is physicalism, according to which colour is definable in the language of physics, for instances as some wavelength-related property.

Clearly, primary quality theories need elaboration if they are adequately to accommodate the evidentiary status of color-experience. Consider in particular the color-knowledge incorporated in CK (2) - (5) above. If no more is said about color than that it is a physical property related to wavelength, it is a mystery how we could have Cartesian certainty about it. Wavelength-related properties are totally non-immanent in experience.

6. A Projective Semantics for Colour Experience

As we have just seen, theories that treat colour as pertaining to subjective experience fail to account for the certainties of colour experience, as do also those that claim that it is an external reality. This accords with Gibson's insistence that sense-features are neither purely subjective nor purely objective, but in some way both. I shall propound a point of view that is in line with Gibson's attitude. But first I have to say something about the *form* of colour experience – the logical form of what it tells us about the outside world.

I want to propose first of all that colour experiences *denote* colours. Denotation is a *semantic* relationship – a relationship that symbols bear to the world in virtue of their *meaning*. Just as the *word* 'circular' denotes the property that circles share, so also Trich's *O*-experience denotes, or represents, a property that orange things share. This gives colour-experience a role quite different from that envisaged by secondary quality theories. We treat sensory experience as a *symbol* internal to the workings of the mind, a token by which the colour vision system passes to other epistemic faculties, and to the perceiver herself, the message that the colour of this visual object is orange. (Remember that the message itself may be true or false – truth and falsity are also semantic relations.) The *O*-experience is, I am suggesting, the semantic marker of this message – semantic not (of course) in the language that the perceiver herself uses (such as English or Malayalam), but in the signalling system that is used by the perceiver's cognitive apparatus.

A semantic theory relies on a mind-world relationship fundamentally different from those employed by the other theories we have been considering. In some of these theories, perceivers have to know something outside the mind before they can come to know colour: in Dispositionalism, perceivers have to know the starting point of a causal relation that (in normal circumstances) vectors inward from world to colour-experience; in the case of Primary Quality Theories, they have to know a physical quantity. Semantic theories, on the other hand, rely on an outbound relation between symbol and denotation. One grasps something of what a symbol signifies when one knows its sense, or meaning. This is what semantic theories require of colour-experience. But since it is possible to grasp the sense of a symbol without being able to identify its referent or denotation, the semantic theory does not require knowledge of anything outside the mind.

The second part of my proposal is that colour experience denotes *projectively*. A projective symbol is one that uses one of its own properties as a proxy for what it denotes. Consider an air-traffic controller working at a radar display. A particular aircraft on her screen is shown as a red dot. She refers to this aircraft as "the red plane". Of course, she does not mean or imply that the aircraft itself is red. She

has no idea what colour it is. She is using a property of the symbol to refer to the object that it denotes. I will call this use of ‘red’ *projective*.

Some projective symbols form a structured denotational system. Think of a map with intervals between equally spaced vertical grid lines marked A, B, C, etc. from left to right and intervals between equally spaced horizontal lines marked 1, 2, 3, etc. from top to bottom. Let us stipulate that this is a projective system in which “Pemberton Street is in G3” means that the Pemberton Street (the external world object) is in the geographic region denoted by grid location G3 (and *not* that the map-representation of the street is in that grid location on the map). Here, location is projectively denoted by the denotative system formed by the symbols on the map.

Certain facts about location are implied by the structure of this projective scheme. Consider:

LK (10) C3 is a greater distance away from A1 than A3.

As stipulated in the preceding paragraph, (10) is about locations in the real space represented by the map in question – it is *location* knowledge, not *map* knowledge. Yet, we know that (10) is true, not in virtue of having measured the distance, but *a priori* as a result of knowing how the denotational system works (and also Pythagoras’ Theorem). (10) is known *a priori*, but it is about the external world.

Notice that (10) is more certain than:

OK (11) Pemberton Street is in G3.

(11) requires not only knowledge of the denotational system, but also empirical information about Pemberton Street. The difference between (10) and (11) is reminiscent of the difference between object-knowledge proposition OK (1) and colour-knowledge propositions CK (2) – (5).

Putting these claims together, we have:

Colour-experiences constitute a structured projective denotational system.

Colour experience is organized around three axes: bright-light, red-green, and blue-yellow. Every colour experience is a combination of values of each of these axes. These experiences are arrayed around these axes as a similarity ordering: the more similar two experiences are, the closer together they are in this system. The qualities of these experiences are projected on to what they denote. To say that something is yellow is to say that it has the colour denoted by the experience we recognize as of the yellow type.

On this view, some colour-knowledge is projected from our knowledge of what it is like to experience colour, and from the projective topology of colour experience. The *colour-knowledge* contained in (2) – (5) is accounted for by innate knowledge of the denotative space of colour; so also our knowledge of missing shades. For instance, CK (2) above assigns a somewhat determinate colour – the colour of the fruit – to an extensive region within Trich’s colour-similarity space – the yellowish region. Her knowledge is analogous to location-knowledge of the following sort: “C3 is contained in the regions between C2 and C4”, which is about location, but known *a priori* as a consequence of knowing the denotational system of grid location. Trich’s knowledge of CK (2) is implicit in her very perception of the fruit as orange. This perceptual state cannot be wrong about the relation stated by (2), though it might be mistaken with respect to the fruit. Propositions like (2) are “quasi-analytic”: all that is needed to grasp them is knowledge of the denotative scheme – and this, I will propose, is innate. Despite their quasi-analyticity, such propositions are “substantive” – they are about relations among the colours.³

Object knowledge as in (1) is explained by general facts about how we use our senses to probe the world. When we look at something in different and demanding ways, our senses converge on a determination that is immune from all but sceptical doubt – i.e., from doubt that does not infect unrelated propositions about our own position or about the condition of our sensory apparatus. Our knowledge of

³ What I am calling “quasi-analytic” is what Kant called the “synthetic a priori”. In Kant’s scheme of things, spatio-temporal experience is a structured projective system. However, he seems to have missed the idea that such systems can be denotative. This is why he is an idealist about space and time.

object-colour depends on determining occurrent facts; this is why it is inherently subject to sceptical doubt.

7. Colour and Epistemic Action

Now the proposal just floated is not yet sufficient to understand our knowledge of colour. For if similarity is wholly dependent on similarity of colour experience, then the proposal becomes a species of subjectivity. Consider the air-traffic controller again. Suppose she has three aircraft on her screen, marked by a red, an orange, and a green dot respectively. The controller can judge that the red dot is more similar with regard to its colour to the orange dot than it is to the green dot. Suppose she projects this judgement onto the aircraft represented by the dots. Then the statement she makes – “The red aircraft is more similar to the orange than to the green” – does not demand of her that she probe the state of affairs beyond her screen. Taken in this way, then, her knowledge of object-colour would be completely immanent – not limitedly immanent. On this construal, she would therefore be entitled to a much higher level of certainty than we have allotted her. So to properly account for the kinds of colour knowledge we mentioned earlier, we need a more robust connection between symbol and denoted object.

This is where epistemic affordance enters the picture. We have attempted to define colour in terms of similarity. But what does colour similarity mean to the perceiver? My proposal is that it denotes a system of epistemic affordances. Sensory similarity relations are the basis of epistemic actions such as conditioning. Let's suppose that feature F is associated with feature G . Given certain other conditions about the naturalness of F and G , an animal that has been exposed to this association will act in the presence of G as if F were presented. In other words, the subjective probability that the animal attaches to F rises when it senses G – its response is conditioned by the association of G with F . It is important to note that the conditioning function – the function that determines how incoming data modifies responses – is itself an unconditioned response. That the animal adopts a conditioned response to F when it finds F associated with G is itself an unconditioned, or innate, response to the association. The perception of F together with G results in a modification of the animal's *epistemic state*.

Now, what happens when G' , which is slightly different from G , is presented? The answer is that the rise in the subjective probability in F will depend on how similar G' is to G . The more similar it is, the greater the probability that the response conditioned on F will be triggered. In fact, this is how similarity space is measured in animals: G is similar to G' to the degree that conditioned responses to G pass over to G' . I shall take this as a definition of similarity – call it similarity-for-conditioning. Thus:

Similarity-for-conditioning: For any animal, x , G is similar to G' to the degree that x 's unconditioned responses to G would overlap, over a long series of trials, with its unconditioned responses to G' .

Here, I should emphasize again that forming a conditioned response – that is, the act of forming such a response – is itself an unconditioned response. The latter unconditioned response is *dynamic*: it is a change in epistemic state (i.e. a change in subjective probabilities) consequent to a perceptual state.

On my proposal, colour-similarity is similarity for the purpose of epistemic actions such as conditioning. Our knowledge of colour is grounded in certain instinctive actions by which we explore and keep records about our surroundings. Those actions are immanent because they are *our* actions; they lack immanence to the extent that their success conditions are determined by things outside ourselves. In fact, colour-similarity has what I have been calling “limited immanence”.

8. Conclusion

Gibson gives sensation an epistemic status all of its own. It is subjective in that it informs us of how the world is apt for certain actions. We know the significance of what sensation tells us precisely because we can measure this significance in terms of these actions. On the other hand, sensation is objective because the senses can misinform us about which actions are called for in a given situation. My

claim has been that when we understand “action” to include not just “behaviour” but also epistemic action, we can gain some perspective on the nature of sensory knowledge.

References

Gibson, James Jerome

(1966) *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

(1979) *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Domain specificity and statistical computations in segmenting fluent speech

Mohinish Shukla
University of Rochester

This paper considers the broad theme of domain specificity and statistical computations in the specific instance of segmenting fluent speech into words. The non-random statistical structure of language can provide important information to a learner. Both adults and infants have been shown to be successful in extracting the distributional structure of controlled artificial ‘languages’ and, further, to analogous stimuli in other perceptual domains. This has led to the proposal that domain-general statistical computations might be sufficient to solve several aspects of the problem of language acquisition. In contrast, it is shown here that even a relatively basic, input-driven computation like speech segmentation is constrained in various, seemingly arbitrary ways, making a truly domain-general computation unlikely. The larger issue of domain-specificity is considered from a neurobiological perspective.

1 The language faculty – innateness and domain specificity

Biological organisms are properly seen in their environmental context, where ‘environmental’ is understood as including the physico-chemical, the biochemical, the ecological and the social aspects. So, a tuna cannot survive outside an appropriate aquatic environment, a predator cannot survive in the absence of prey and, in more social creatures like macaques, social deprivation, specially early in life, leads to decreased survival rates (Lewis et al, 2000). As a corollary, an organism can be seen as a biological entity that embodies the ability to navigate through its own specific environment, from the single-celled zygote to the mature form. In this conception, one can meaningfully ask: what are the specific structures that an organism embodies that allow it to successfully operate in its environment?

In the classical Darwinian view, organisms evolve in response to their environments, such that their bodily structures represent adaptations to their environment (e.g., Dawkins, 1986). Today, we know that the environment, as described above, plays a significant role in determining the mature form of the organism, both evolutionarily speaking and in the lifetime of an individual organism. To be clear, no amount of change in the immediate environment could result in an elephant embryo developing into a fruit fly. Nevertheless, the mature phenotype can be conditioned by the environment. For example, in certain reptiles, the sex of an individual is determined by the temperature of incubation of the egg, and represents an adaptive response to the environment (Warner & Shine, 2008). For Darwin, mental and behavioral propensities were also seen as inherited adaptations to the environment (Darwin, 1859). And indeed, this was the position of early ethologists from the last century (e.g., Tinbergen, 1951). It was also argued that the development of an instinct could not be decoupled from the environment any more than could the growth of the body (e.g., Lehrman, 1953). From this perspective then, for any given trait – physical, behavioral or mental – the contributions of the biology (the genetic underpinnings) of the organism and of the environment in the ontogeny and maturation of the trait is strictly an empirical issue. One might then meaningfully refer to those (innate) aspects that are, to paraphrase Hebb (1949, pg 166, cited in Marler, 2004), “... predictable from the acknowledgment of the species, without knowing the history of the individual animal.”

Human language can be seen as a trait that is predictable from the acknowledgment of the species. The presence of this faculty is a defining feature of our species, and hence must rely in some part on innate predispositions. Language is not a single, monolithic domain, and different parts of language (like syntax, morphology, phonology) might have their own phylogenetic and ontogenic routes (e.g., Jackendoff & Pinker, 2005). However, there is a further distinction to be made: is the faculty for language an independent mental domain, or is this faculty cobbled together from various mental competences,

some unique and some shared with other species? In a sense, these questions parallel those about the contributions of genes and the environment in the development of an organism: the development of the language is seen as the interaction of specifically linguistic competences and a *cognitive environment*, composed of other cognitive systems, along with the social input from other members of the species (see also Hauser, Chomsky & Fitch, 2002).

Why must there be a specifically linguistic competence? In addition to the observation that only humans appear to have this capacity, Chomsky puts forward a second argument. Characterizing language as an abstract, rule-based system operating over symbolic representations (but see, e.g., Elman et al, 1996), he points out that in the absence of explicit teaching of the underlying rule system, it is in theory virtually impossible to infer it from the finite input that a child receives (e.g., Chomsky, 1965). Indeed, this is the familiar problem of induction – generalizing general principles from a limited set of observations. However, this argument does not imply that all mental structures that support the language faculty have to be *specifically linguistic*; such abstract structures might be part of the broader cognitive environment (e.g., Perfors, Tenenbaum & Regier, 2006). Indeed, some linguists have invoked non-linguistic processing constraints in order to describe various aspects of language like morphosyntax (e.g., Comrie, 1981; Hawkins, 1988; DuBois, 1987; Cutler, Hawkins, & Gilligan, 1985, Fenk-Oczlon and Fenk 2004, amongst others). Similarly, Endress and colleagues (e.g., Endress & Mehler, submitted, Endress & Mehler, in press, Endress, Dehaene-Lambertz & Mehler, 2007) have suggested that language learning and processing relies on perceptual primitives, most of which might be shared at least with non-human primates.

Several authors have appealed to the notion of *domain general computations* in order to understand language acquisition and processing (e.g., Landauer & Dumais, 1997, Reali & Christiansen, 2005, Foraker et al, 2007). In the next section, we look at what domain general might mean.

2 Domain general computations

In a certain sense, the term ‘domain general’ is not much different from the term ‘domain specific’. Consider what it might mean for a computation to be domain general. Let us take a concrete case: Bayesian inference (e.g., in Perfors, Tenenbaum & Regier, 2006 and Foraker et al 2007). Although these authors study the specific instance of learning a linguistic rule for which the data is insufficient, in general terms their domain general computation takes as input a very specific kind of data (linguistic data in their case) and uses this data to infer the most likely of a given set of structures (grammars in this case). For this computation to be truly domain general, it would have to treat data from any domain in a similar way, and use the same inference method to make inferences over the same set of structures. That is, in every case, the treatment of the data, the set of priors¹ and the computation to compute the posteriors would have to be the same. This implies that any data, visual, auditory, tactile etc., would be evaluated in the same way, and two sets of data that share the same statistics should result in the same output. Over the last decade or so, evidence has been accumulating that one particular kind of computation might indeed be domain general in this sense.

Saffran, Aslin & Newport (1996, henceforth SA&N) considered the problem of segmenting speech into a series of words. Although most theories of syntax acquisition treat words (or morphemes) as the basic input, words are not immediately apparent in fluent speech as might be intuitively obvious upon listening to speech in a foreign language. Words are not reliably marked, acoustically or otherwise, and thus the first task of the infant learner would be to identify the words of the language (but see Nespor et al, 2007, and Shukla & Nespor, in press, for an alternative). SA&N asked if infants might, in part, rely on a general computational strategy for segmenting fluent speech. They relied on the observation that, in general, the syllables that make up a word are expected to be more (statistically) coherent than syllables across a word boundary. To get an intuitive idea, consider a word like ‘pretty’. The first syllable is shared by other words (like ‘prickle’ or ‘primitive’), so hearing just the first syllable, we might have some idea of

¹ In Bayesian inference, *priors* refer roughly to the set of prior hypotheses (e.g. the set of all possible structures) along with their relative probabilities in the absence of any observations. The *posterior* is the re-evaluated likelihoods after having observed the data. The data is used to evaluate which of the prior hypotheses is most likely given both the probability of the hypothesis itself and the probability of observing the data given that hypothesis.

the possible words. However, upon hearing the second (final) syllable, what can we expect as the next syllable? Clearly, it could be the first syllable of any word that can follow the word ‘pretty’. That is, while there is some predictability from the first syllable of ‘pretty’ to the second, the predictability from the second to the next is much poorer. SA&N formalized this intuition as the *forward transition probability* (TP for short):

$$(1) TP(x \rightarrow y) = \frac{frequency(x,y)}{frequency(x)}$$

The TP can be thought of as the probability of the next syllable being y , given that the current syllable is x . In other words, it is the predictability of the upcoming syllable given the current one – the higher the TP between x and y , the more likely is y to follow x . Conversely, if the TP from x to y is low, it would indicate low predictability, and hence a possible word boundary.

Saffran and colleagues demonstrated that both infants and adults could use TPs to segment controlled, artificial speech stimuli, in which TPs were the primary or the only cue to ‘word’ boundaries (e.g., Saffran, Aslin & Newport, 1996, Saffran, Newport & Aslin, 1996, Aslin, Saffran & Newport, 1998, Saffran, 2001, Peña et al, 2002, Thiessen & Saffran, 2003). Subsequently, it was demonstrated that one could replace the syllabic unit with units in other perceptual domains and observe similar segmentation results. Thus, these experiments were extended to various auditory ‘units’ including tones (Saffran et al, 1999, Creel, Newport & Aslin, 2004), complex sound patterns (Gebhart, Newport & Aslin, 2009), timbre (Tillman & McAdams, 2004), patterned visual elements (Fiser & Aslin, 2002, Kirkham, Slemmer & Johnson, 2002) and even patterns of motor movement (Hunt & Aslin, 2002). These findings might be taken to indicate that there is a domain-general mechanism for computing TPs (see, e.g., Kirkham, Slemmer & Johnson, 2002). We will examine just the segmentation of speech in Section 4, but first let us consider the nature of a domain general computation from a neurobiological perspective.

3 A neocortical perspective on domain generality

As any basic textbook on mammalian neuroanatomy will show, inputs from the different sensory modalities take different paths and arrive at fairly well defined cortical fields, known as the primary sensory cortices. These primary sensory cortices vary between species and to some extent reflect differences in their ecological adaptations (Kaas, 1989, Krubitzer, 2007), although they share common organizational themes (Krubitzer, 1995). However, the fact that different sensory modalities project to different cortical areas raises the question: how can there be a (domain general) central processor that acts equally on any kind of input?

One possibility is that primary cortices might act as way stations, relaying appropriately coded information to the central processes located elsewhere, as in association cortices, which are known to be multimodal. Nevertheless, even association cortices do not necessarily always receive input from all the sensory modalities. Indeed, the role of association cortices appears to be more of multimodal *integration*, rather than as common processing stations for input from any modality (e.g., Kaas & Collins, 2004).

More importantly, different contexts might call for a certain computation to be carried out over different “units” in the input. For example, speech can be considered as a sequence of syllables, but it can also be considered as a sequence of phonemes or of phrases, and it is not clear how the appropriate unit might be selected (see Section 4 for a further discussion on this point).

A second way in which domain general computations might be implemented in the brain is if the cortex itself is homogenous, such that all the cortical areas are structured in the same way. We know that the adult cortex is structurally heterogenous; for example, this heterogeneity is the basis for subdividing the cortex into the various Brodmann areas. However, it is not clear whether these differences are innate or are driven by differences in the patterned activity received from different sensory sources. Krubitzer & Kahn (2003) examine the developmental molecular neurobiological aspect of this question, and find that different aspects and different areas rely differently on genetic and experiential factors. More recently, Sur and colleagues have demonstrated in a series of elegant experiments that functionally, one patch of cortex can carry out some of the functions of another (e.g., Angelucci, Clasca & Sur, 1998, Sharma, Angelucci & Sur, 2000). In their ferret re-wiring experiments, by careful surgical procedures in newborn ferrets, these authors were able to redirect visual input from the eyes to the primary auditory cortex (the

visual cortex received no input). These authors observed that the rewired auditory cortex showed many neuronal organizational characteristics of the normal primary visual cortex, and some like the normal primary auditory cortex. Importantly, behaviorally these animals were able to see, although their vision was poorer than normal animals (von Melchner, Pallas & Sur, 2000)².

These studies suggest that, in some respects, the cortical tissue might indeed be functionally similar in different areas of the brain. Therefore, it is possible that there are some computations that are similar across different modalities and different stimuli. Nevertheless, it is also clear that there are differences across cortical areas, and thus some computations (or some aspects of some computations) might be domain specific, and hence different across modalities.

In this view, the mind/brain is seen as being an organ pre-wired to solve certain tasks in a variety of domains, like depth perception, cheater detection or language acquisition. The task thus determines the kind of information most relevant to it, and statistical computations are one way of gathering such information. That is, similar statistical computations of any sort are not seen as arising from shared resources at a common, central processor, but are seen as the tools used by different cortical areas in solving different tasks specific to those areas. They could be similar across domains either through similarities in cortical structure and function, as described above, or might have arisen independently to solve separate tasks that required that common computation.

As described above, it is not clear how a general statistical computation might pick the right ‘units’ over which to generate answers. However, this is the least of the problems with a domain general approach. In the next section, we will look at the specific instance of using TPs to segment fluent speech in greater detail. We will see that this task is constrained and biased in several ways. That is, the task of finding the appropriate units is so complex and multi-faceted that even if there were a central processor, it will be clear that the bulk of the work must be done by specific modules that can gather just the right kinds of input as data (see also Yang, 2004).

4 Speech segmentation

As described in Section 2, words are not clearly marked in fluent speech. Nevertheless, there are statistical regularities that could be exploited to find word boundaries (see Charniak, 1993). For example, the sequence ‘k-n’ is rare inside words in English, but not uncommon across word boundaries. Therefore, positing a word boundary between a ‘k’ and an ‘n’ will on average be a fairly successful segmentation strategy in English (e.g., Church, 1987). There have thus been several proposals for generic strategies that rely on statistical computations over sub-lexical units (e.g., phonemes, syllables etc) to discover word boundaries (e.g., Harris, 1995, Brent & Cartwright, 1996, Gow, Melvold & Manuel, 1996, Dahan & Brent, 1999, Batchelder, 2002). However, such strategies by themselves might not be enough to solve the problem of word segmentation (see also Yang, 2004). In this section, we will look at various constraints and biases that have been shown to influence segmentation of words even in simple, artificially controlled stimuli.

To begin, at a very fundamental level, it is clear that words represent a certain, specific coding strategy. To see that words are coded in a specific way, let us compare it to another coding scheme that does things a little differently; let us consider an analogy between words as portions of speech utterances that encode a specific meaning and genes as portions of a DNA strand that encode a specific protein (Lewin, 2008)³. Now, while it is true that some genes are continuous portions of DNA just as words are continuous portions of the speech stream, there are also significant differences. For example, genes need not be continuous on the DNA, but can be interrupted by other coding or non-coding regions of DNA. This would be equivalent to having a sequence ‘gui.he.tar.ro’⁴ stand for ‘guitar hero’. Second, genes can overlap, such that some bases are shared between two adjacent genes. This would be equivalent to having

² Similar structure alone is not sufficient to warrant the claim of similar function. For example, while it is widely believed that the organization of monocular eye cells into interdigitating stripes have a functional significance, new findings are challenging this proposal (see Horton & Adams, 2005, for an overview).

³ DNA, like speech, has a directionality, so the situation is indeed quite analogous to the case of speech.

⁴ Periods mark syllable boundaries.

a sequence ‘fif.teen.agers’ stand for ‘fifteen teenagers’. Clearly, words don’t have either of these properties; they cannot be discontinuous, nor overlapping.

In general, one could come up with a host of different possible coding schemes. The relevant point is that words are not simply any kind of coding scheme, but a very specific one. A truly domain general solution would need to be able to determine *any* such coding scheme. Indeed, recent work by Tenenbaum and colleagues (e.g. Kemp & Tenenbaum, 2008) propose such a mechanism. In their view, therefore, the innate component might be just an algorithm of that sort, with even the coding scheme being inferred from the input. However, even if it is in theory possible to determine the appropriate coding scheme from the input, we will still need to explain the various constraints and biases, some of which are discussed below.

Before discussing the constraints and biases, let us look at the kind of experimental paradigms that researchers have used to address the speech segmentation problem. In the simplest design, the experimenter creates a set of nonce words (typically around 6), like *pu.ra.ki* or *be.fo.du*. The nonce words are then concatenated into a continuous stream, where all the nonce words are repeated several times in random order. There are no pauses or any other prosodic cues to the word boundaries, and the onsets and offsets of such speech ‘streams’ are ramped up and down in amplitude so the listener cannot perceive a beginning or an end. Participants are exposed to such speech streams, and subsequently tested for their recall or preference for a nonce word over a *part-word* – a sequence of syllables made up from parts of one word and parts of another. For example, the nonce word *pu.ra.ki* might be pitted against a part-word *ra.ki.be*. Adult participants find nonce words significantly more familiar than the part-words, or rate them as more likely to have been heard, or to belong to the language they were familiarized with. Infants show a discrimination of the two kinds of sequences by looking longer when one kind of sequence is played, compared to the other; typically displaying a *novelty preference*, wherein they look significantly longer for the part-words.

Using such a paradigm, researchers have found that the units over which TPs are computed play an important role. While the syllable has been long considered a fundamental unit of speech (e.g., Mehler 1981, Bertoncini & Mehler, 1981), it has been shown by Bonatti and colleagues (Bonatti et al, 2005, Mehler et al, 2006) that, if syllable TPs are held constant and nonce words are defined over the consonants alone, then segmentation is still possible. However, when the nonce words are defined over the vowels alone, they are not segmented (but see below, and Toro et al, 2008). Thus, the first constraint is about the choice of the units: consonants (and syllables) are preferred over vowels.

In addition to probing the units of segmentation, researchers have also asked: what is that nature of the computation involved? Several investigators have considered forward TPs, backward TPs, mutual information, clustering, co-occurrence frequency and Bayesian inference (e.g., Brent & Cartwright, 1996, Christiansen, Allen & Seidenberg, 1998, Perruchet & Vintner, 1998, Swingley, 2005, Orban et al, 2008). In Shukla (2007) it was found that increasing the complexity of the task by inserting a large number of random syllables that lacked any statistical structure did not seem to interfere with the extraction of embedded nonce words that had high TPs between their constituent syllables (see also Frank, Gibson & Tenenbaum, 2009). However, if a nonce word sometimes occurred in close proximity to itself, it was better segmented than another nonce word that never occurred in close proximity to itself, suggesting a memory constraint on such computations.

Thus, constraints are also placed by more cognitively general resource like memory. This raises serious concerns about, for example, a Bayesian model, in which such proximity (memory) effects are not expected in a straightforward way. Further, even attention, another common cognitive resource, can modulate TP computations both in the auditory (speech, Toro, Sinnett and Soto-Faraco, 2005) and the visual domain (Baker, Olson & Behrmann, 2004 and Turk-Browne, Jungé & Scholl, 2005).

Further, Endress and colleagues have shown (Endress & Mehler, in press) that TP computations have a problem with transitivity – if the TP from syllable A to syllable B is high, and the TP from syllable B to syllable C is high, then the sequence ABC should be a good ‘word’ candidate even if it has never occurred. And indeed, this is what adult participants (mis)perceive. Endress et al therefore conclude that TPs cannot be a solution to the word learning problem, but at best provide biases that must be confirmed by other means before a high-TP sequence can be considered a real word.

Next, several researchers have shown that such TP computations are greatly affected by (or rely upon) perceptual phenomena. For example, Creel, Newport & Aslin (2004) found that perceptual

similarity constrained how the TP computations were made in the auditory domain. In particular, TPs over non-adjacent tones were computed only when they were perceptually similar, leading to a ‘streaming’ effect. Similarly, Fiser, Scholl & Aslin (2007) showed that perceptual grouping in the visual domain influences visual statistical learning.

Perceptual salience also plays a role in TP computations by highlighting certain units over others. For example, Shukla (2006) found that sharp changes in pitch (pitch ‘breaks’) that serve to carve fluent speech into a series of ‘phrases’ also serve to highlight syllables at the edges of such breaks; nonce words formed from these syllables are better extracted than statistically identical nonce words at non-salient locations. Immediate repetitions are known to be perceptually salient. Indeed, in most of these segmentation experiments, immediate repetitions are disallowed since they are immediately extracted from the artificial speech streams (M. Peña & J. Mehler, pers. comm.). In fact, if repetitions are allowed over the vocalic tier, then, contra Bonatti et al (2005), TPs can be computed even over the vocalic tier (Newport & Aslin, 2004, Mehler et al, 2006), suggesting that the units of computation (consonants or vowels) might themselves be conditioned by other perceptual factors.

Finally, TP computations can be affected by prosody. Fluent speech is organized into prosodic domains ranging from syllables to phrases and entire utterances. These domains are marked (with some variation between languages) by acoustic cues like changes in pitch and the duration of various segments. These cues serve to mark prosodic domains, and, since words are aligned with larger prosodic phrases, such cues can be useful in identifying word boundaries. In most of the experiments described above, the speech streams are created such that they lack any prosodic features, so investigators can determine if solely statistical cues can be utilized to segment speech. In Shukla, Nespor & Mehler (2007) we examined the effect of prosody on such statistical computations (see also Shukla, 2006). To summarize the results, we found that prosody did not appear to restrict the domain over which statistical computations were carried out. That is, when probed for their memory of high-TP syllable sequences, participants showed evidence of remembering all such sequences, whether they were prosodically appropriate or not. However, only the prosodically appropriate sequences were treated as possible words in the artificial language. That is, we found a dissociation between merely remembering syllabic sequences and treating them as possible words. Indeed, under certain circumstances, participants preferred to treat sequences they had never heard before as possible words, over high-TP but prosodically inappropriate sequences.

More recently, we also found evidence for a case in which prosody does appear to actually restrict the domain of computation (Shukla & Nespor, 2008). In many languages of the world, the vowels inside a (prosodic) word tend to become more similar to each other (harmonize). We thus asked if sequences of syllables with *identical* vowels would be preferred as words. We tested adult Italian participants who, in the absence of any familiarization, have no preference for syllable sequences with or without identical vowels. We then familiarized a different group of Italian participants with artificial speech that consisted of trisyllabic words defined by high TPs over the consonants. However, the stream was so constructed that the part-words, although they had lower TPs over the consonants, always had the same vowel. Following familiarization with such stimuli, participants now showed a significant preference for the statistically less coherent part-words. Further, we found little evidence to suggest that they even considered or extracted the high-TP sequences.

Finally, we are yet to explore the role of social constraints even for tasks like word segmentation. For example, we know that there is significant speaker variation. Does this mean that the infant must track the statistics between syllables for all the speakers they encounter? What aspects of speaker variations are retained and what (if any) are thrown away? As Saussure (1983) suggested more than a hundred years ago, a linguistic contrast is only meaningful if the individual chooses to make a distinction. That is, variability within a speaker must somehow be disentangled from variability across speakers.

5 Conclusions

The assumption of a domain general learning mechanisms is, to cite Gallistel (1999), “...equivalent to assuming that there is a general purpose sensory organ, which solves the problem of sensing.” Indeed, it is hard to see what a domain general perspective can offer for a task like segmenting words from fluent speech. In this article, it is suggested that the brain does not merely compute statistics from the input it receives using a monolithic, central statistical processor, but instead uses various sources

of information in order to solve pre-defined tasks. This information can come from highly specialized structures that mediate its rapid acquisition, and might also rely on general constraints placed by the various aspects of the environment as described in Section 1. Ultimately, this is an empirical question.

So how can we account for the so-called domain general computations? In one sense, the domain generality of a computation (for acquisition or otherwise) might arise from shared neural function (with or without shared architecture, Section 3) between cortical areas that process different kinds of input. However, there is a second sense implicit in the discussion about perceptual processes or constraints from other cognitive domains like memory (Section 4): since the different sensory modalities all live in the same brain, they might be subject to similar constraints from the cognitive environment (see also Sperber, 2004).

References

- Angelucci A., Clasca, F., & Sur, M. (1998) Brainstem inputs to the ferret medial geniculate nucleus and the effect of early deafferentation on novel retinal projections to the auditory thalamus. *J. Comp. Neurol.*, 400, 417–439
- Aslin, R.N., Saffran, J.R., & Newport, E.L. (1998). Computation of conditional probability statistics by human infants. *Psychological Science*, 9, 321-324
- Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science*, 15, 460-466.
- Batchelder, E. (2002). Bootstrapping the lexicon: A computational model of infant speech segmentation. *Cognition*, 83 (2), 167-206.
- Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant speech perception. *Infant Behavior and Development*, 4, 247-260.
- Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on Statistical Computations. *Psychological Science*, 16(6):451-9.
- Brent, M., & Cartwright, T. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, 61(1-2):93-125.
- Charniak, E. (1993). *Statistical language learning*. Cambridge, MA: The MIT Press.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA : The MIT Press
- Christiansen, M., Allen, J., & Seidenberg, M. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, 13, 221-268.
- Church, K. (1987). *Phonological parsing in speech recognition*. Dordrecht: Kluwer Academic.
- Comrie, B. (1981). *Language universals and linguistic typology*. Basil Blackwell.
- Creel, S. C., Newport, E. L., and Aslin, R. N. (2004). Distant melodies: Statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1119-1130.
- Cutler, A., Hawkins, J.A., and Gilligan, G. (1985) The suffixing preference: a processing explanation. *Linguistics*, 23, 723-758
- Dahan, D., & Brent, M. (1999). On the discovery of novel word like units from utterances: An artificial-language study with implications for native-language acquisition. *J Exp Psychol Gen*, 128(2):165-185.
- Darwin, C. (1859) Instinct. In *The Origin of Species by Means of Natural Selection or The Preservation of Favoured Races in the Struggle for Life. Chapter 7*. London: John Murray. 234-263.
- Dawkins, R. (1986) *The blind watchmaker*. New York: W.W. Norton & Co.
- DuBois, J. (1987). The discourse basis of ergativity. *Language*, 64, 805–855.
- Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: The MIT Press.
- Endress, A.D., Dehaene-Lambertz, G., & Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition*, 105(3):577-614.
- Endress, A.D. & Mehler, J. (submitted). Primitive Computations in Speech Processing.
- Endress, A.D. & Mehler, J. (in press). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language*.

- Fenk-Oczlon, G. & Fenk, A. (2004). Systemic Typology and Crosslinguistic Regularities. In: V. Solovyev & V. Polyakov (eds.) *Text Processing and Cognitive Technologies*, 229-234. Moscow: MIS.
- Fiser, J., & Aslin, R.N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99, 15822-15826.
- Fiser, J., Scholl, B. J., & Aslin, R. N. (2007). Perceived object trajectories during occlusion constrain visual statistical learning. *Psychological Bulletin and Review*, 14, 173-178
- Foraker, S., Regier, T., Khetarpal, N., Perfors, A., & Tenenbaum, J.B. (2007) Indirect evidence and the poverty of the stimulus: The case of anaphoric one. *Proceedings of the 29th Annual Conference of the Cognitive Science Society*.
- Frank, M., Gibson, E., & Tenenbaum, J (2009) *Large-Scale Statistical Segmentation With Naturally-Distributed Word Frequencies* Talk presented at SRCD 2009 Biennial Meeting, Denver.
- Gallistel, C. R. (1999). The replacement of general-purpose learning models with adaptively specialized learning modules. In M.S. Gazzaniga, (ed.). *The Cognitive Neurosciences. 2d ed.* (1179-1191) Cambridge, MA: The MIT Press
- Gebhart, A. L., Newport, E. L., and Aslin, R. N. (2009). Statistical learning of adjacent and non-adjacent dependencies among non-linguistic sounds. *Psychonomic Bulletin & Review*, 16, 486-490.
- Gow, D., Melvold, J., & Manuel, S. (1996). How word onsets drive lexical access and segmentation: evidence from acoustics, phonology and processing. *Proc. ICSLP96*.
- Harris, Z. (1955). From phoneme to morpheme. *Language*, 31, 190-222.
- Hawkins, J. (1988). Explaining language universals. In J. Hawkins (Ed.), *Explaining language universals*. Basil Blackwell.
- Hauser, M.D., Chomsky, N., & Fitch, W.T. (2002) The faculty of language: what is it, who has it, and how did it evolve?. *Science*, 298(5598):1569-79.
- Hebb, D.O. (1949) *The organization of behavior: A neuropsychological theory*. New York: John Wiley & Sons.
- Hunt, R.H., & Aslin, R.N. (2001). Statistical learning in a serial reaction time task: Simultaneous extraction of multiple statistics. *Journal of Experimental Psychology: General*, 130(4), 658-680
- Jackendoff, R., & Pinker, S. (2005) The nature of the language faculty and its implication for evolution of language (Reply to Fitch, Hauser and Chomsky). *Cognition*, 97, 211-225
- Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, 105(31):10687-10692
- Kaas, JH (1989) The evolution of complex sensory systems in mammals. *J Exp Biol* 146, 165-176
- Kaas, J.H. and Collins, C.E. (2004) The resurrection of multisensory cortex in primates: connection patterns that integrate modalities. In: G.A. Culvert, C. Spence and B.E. Stein (Eds) *The Handbook of Multisensory Processes*, Cambridge, MA: The MIT Press, 285-293.
- Kirkham, N.Z.; Slemmer, J.A.; Johnson, S.P. (2002) Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, 83, B35-B42
- Krubitzer, L (1995) The organization of neocortex in mammals: are species differences really so different? *Trends in Neurosciences*, 18(9):408-417
- Krubitzer, L (2007) The magnificent compromise: cortical field evolution in mammals. *Neuron*, 56, 201 – 208
- Krubitzer, L & Kahn, DM (2003) Nature versus nurture revisited: an old idea with a new twist. *Progress in Neurobiology*, 70, 33-52
- Landauer, T., & Dumais, S. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211-240
- Lehrman DS. 1953. A critique of Konrad Lorenz's theory of instinctive behavior. *Quart Rev Biol*, 28, 337-363
- Lewin, B. (2008) *Genes IX*. Jones & Bartlett
- Lewis, M.H., Gluck, J.P., Petitto, J.M., Hensley, L.L., & Ozer, H. (2000) Early social deprivation in nonhuman primates: long-term effects on survival and cell-mediated immunity. *Biol Psychiatry*, 47(2):119-126

- Marler, P. (2004) Innateness and the instinct to learn. *Anais da Academia Brasileira de Ciências*, 76, 189-200.
- Mehler, J. (1981). The role of syllables in speech processing: Infant and adult data. *Philosophical Transactions of the Royal Society*, 295, 333-352.
- Mehler, J., Peña, M., Nespors, M. & Bonatti, L. L. (2006). *The "Soul" of language does not use statistics: Reflections on Vowels and Consonants*. *Cortex*, 42, 846-54
- Nespors, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008) Different phrasal prominence realizations in VO and OV languages? *Lingue e linguaggio*, VII.2: 1-28.
- Orban, G., Fiser, J., Aslin, R. N., and Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, 105, 2745-2750.
- Peña, M., Bonatti, L., Nespors, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298(5593):604-607.
- Perfors, A., Tenenbaum, J., Regier, T. (2006) Poverty of the Stimulus? A rational approach. *28th Annual Conference of the Cognitive Science Society*. Vancouver, British Columbia.
- Perruchet, P., & Vinter, A. (1998). Parser: A model for word segmentation. *Journal of Memory and Language*, 39, 246-263.
- Real, F. & Christiansen, M. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science*, 29, 1007-1028
- Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, 81(2):149-169.
- Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294):1926-1928.
- Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by adults and infants. *Cognition*, 70, 27-52.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606-621.
- Saussure, F. de. (1983 translation by Roy Harris) *Course in General Linguistics*. La Salle, Illinois: Open Court.
- Sharma J., Angelucci, A., & Sur, M. (2000) Induction of Visual Orientation Modules in Auditory Cortex; *Nature (London)* 404, 841-847
- Shukla, M. (2006) Prosodic constraints on statistical strategies in segmenting fluent speech. Unpublished Ph.D. Dissertation, SISSA, Trieste.
- Shukla, M., and Nespors, M. (2008) *The vowel tier constrains statistical computations over the consonantal tier*. Poster presented at AMLaP, September 2008, Cambridge, UK.
- Shukla, M. & Nespors, M. (in press) Rhythmic patterns cue word order In: L. Rochman, & N. Erteschik-Shir (Eds.) *The Sound Pattern of Syntax*. OUP: Oxford.
- Shukla, M., Nespors, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, 54, 1-32.
- Sperber, D. (2004). Modularity and relevance: How can a massively modular mind be flexible and context-sensitive? in P. Carruthers, S. Laurence & S. Stich (Eds.), *The Innate Mind: Structure and Content*. Oxford University Press
- Swingle, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, 50, 86-132.
- Thiessen, E., & Saffran, J. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Dev Psychol*, 39(4):706-716.
- Tillman, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *JEP: LMC* 30(5):1131-1142
- Tinbergen, N. (1951) *The study of instinct*. Oxford: Clarendon Press.
- von Melchner L., Pallas, S.L., & Sur, M. (2000) Visual Behaviour Mediated by Retinal Projections Directed to the Auditory Pathway; *Nature (London)*, 404, 871-876
- Toro, J.M., Shukla, M., Nespors, M., & Endress, A.D. (2008). The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences. *Perception and Psychophysics*, 70(8):1515-1525.

- Toro, J.M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, 97, B25-B34.
- Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, 134, 552-564
- Warner, D.A., & Shine, R. (2008) The adaptive significance of temperature-dependent sex determination in a reptile. *Nature*, 451(7178):566-568.
- Yang, C. D. (2004) Universal Grammar, statistics or both? *Trends in Cognitive Sciences*, 8(10):451-456

Mohinish Shukla
mohinish.s@gmail.com

Public and private patterns in language

Diana Archangeli
University of Arizona

What is the role of an innate linguistic capacity (Universal Grammar), and how much emerges as the human brain processes linguistic information? I address this question through attention to the distinctive features of sounds and an examination of phonological (nasal place assimilation), allophonic (/s/ retraction), and idiophonic (pronunciation of /ɹ/) patterns in American English. Regardless of the type of sound and sound pattern under consideration, the same tasks of categorization, classification, and organization are required during language acquisition. Emergence presents a unified view while Universal Grammar requires two distinct mechanisms, one for patterns involving distinctive features and one for other types of patterns. However, the mechanism necessary for the allophonic and idiophonic phonetic patterns does everything that is needed for the phonological patterns as well. Thus there is unneeded formal duplication under the Innatist hypothesis.

1 Introduction

The question of what is innate and what emerges directly from the data is central in understanding how the human mind works. So too is the study of how language is structured. In this paper, I consider the question of innateness vs. emergence in the domain of language structure, specifically with respect to language sounds.

Sounds are used in different ways in language. They are used *phonemically* to make a meaningful difference between words, e.g. the final nasals in [kæm] ‘cam’ vs. [kæn] ‘can’. They are also used *allophonically* to make words “sound right”, even though the difference is not meaningful, such as the contrast between the the pronunciation of “t” in *atom* [æɾəm] vs. *atomic* [ətʰəmɪk]. In this paper, we examine Emergence vs. Innatist implications for phonemic and allophonic patterns, as well as a third type of pattern, those peculiar to an individual, or *idiophonic* patterns.

1.1 Distinctive features

Distinctive features are used to account for sound patterns in languages, see Trubetzkoy (1936); Jakobson *et al.* (1952); Chomsky & Halle (1968). Distinctive features are used to describe individual sounds and to classify sounds as being similar to or different from each other (by the number and type of shared features). Distinctive features also allow comparison across languages by classifying the sounds of each language with the same terms. As more languages are studied, generalizations across languages (universal patterns) are expressed in terms of distinctive features. Additionally, distinctive features are used to characterize language sound patterns.

Distinctive features also offer a point of contrast between Emergence and Innatist hypotheses. Under the Emergence hypothesis, distinctive features emerge during the acquisition of a language; which features emerge depends on properties of the language being acquired. Under the Innatist view, acquisition includes mapping innate distinctive features to the sounds encountered in the language being acquired.

1.2 Innate vs. Emergent

There are some advantages of innate distinctive features. For example, innate distinctive features account for the high degree of similarity across the sounds used in languages. If the same set of features is available as part of the genetic make-up of each normal human, then it is expected that languages will use this set of features and so evince a high degree of similarity. However, the human vocal tract, used to produce sounds, is quite similar across humans, as is the human auditory system, used to perceive sounds. The emergence view predicts a high degree of similarity in the linguistic ways that the people use their articulatory and auditory systems. This is exactly what is found. By contrast, the innatist perspective can make no such claim: it is entirely accidental that the genetic endowment for language maps closely to the articulatory and acoustic systems.

Similarly, innate distinctive features allow for the expression of Sound Universals. This becomes more relevant in specific versions of phonological theory. For example, in Optimality Theory, patterns are expressed in terms of strictly ranked universal constraints, some of which refer to features. These universals, then, are universal tendencies, not absolute universals. However, again due to the high similarity among human vocal tracts and auditory systems, it is not surprising that the same constraints recur in a variety of languages. The strong argument for innateness would be robust Sound Universals that do not have articulatory or acoustic explanations. These do not appear to exist.




Finally, innate distinctive features might be viewed as answering the “poverty of the stimulus” argument, making it easier for the language learner to acquire the sounds and sound patterns of his/her language. As shown here, innate distinctive features actually make the job more difficult because there is an additional level of acquisition: figuring out which set of distinctive features to use for the sounds being acquired. See Chomsky & Halle (1968); Prince & Smolensky (1993); McCarthy & Prince (1993) for models adopting Innateness; see Blevins (2004); Mielke (2004, 2005); Mohanan *et al.* (to appear); Port (2007) for more on Emergence.

As an interim conclusion, there is skepticism about the arguments in support of innate distinctive features. Emergence answers those arguments. The next question to ask is whether there is also evidence supporting Emergence, to be found in sound patterns. In the remainder of this paper, we consider three types of sound patterns from English, phonological, allophonic, and idiophonic, and conclude that the Emergence hypothesis makes more sense than the Innatist hypothesis.

2 Phonological patterns

In some cases, certain sounds have an altered pronunciation in certain environments. This is considered a phonological alternation when the resultant sound is a distinct sound of the language. English has three distinct nasal sounds, illustrated in (1). The dorsal nasal /ŋ/ has a limited distribution in that it never occurs at the beginning of a word. However, as (1) shows, /ŋ/ contrasts with both /m/ and /n/ in word-final and word-medial position. (There are other types of nasals in English, for example the dental nasal found in words like ‘tenth’, [tɛnθ]. This is an allophonic variation, not a phonological alternation, because the dental nasal does not play a contrastive role in English.)

(1) **Three English nasals**

LABIAL [m]		CORONAL [n]		DORSAL [ŋ]	
[ɪɹɒm]	‘rum’	[ɪɹɒn]	‘run’	[ɪɹɒŋ]	‘rung’
[sɪməɹ]	‘simmer’	[sɪnəɹ]	‘sinner’	[sɪŋəɹ]	‘singer’
					

English nasals also participate in a phonological alternation: they take on the place of articulation of the following stop, illustrated in (2). What these data show is that exactly the same features both distinguish the nasals from each other and play a significant role in this alternation, LABIAL, CORONAL, and DORSAL.

(2) **The same features are significant in nasal place assimilation**

before LABIAL	[m]	i[mp]ossible	i[mb]alance
before CORONAL	[n]	i[nt]olerant	i[nd]estructible
before DORSAL	[ŋ]	i[ŋk]onclusive	i[ŋg]ratitude

When encountering this pattern, it is critical that the language learner figure out that /m/, /n/, and /ŋ/ are distinct sounds in English; the language learner must figure out that there are three categories of nasals. While this sounds simple, it actually is quite a complex task: nasal percepts must be collected and then evaluated to determine whether there are categories. Utterances of “the same” nasal sounds differ, both within speakers and across speakers, so there are challenges involved.



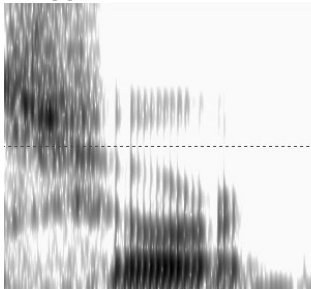
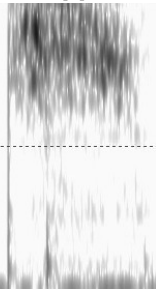
The learner must also figure out how to produce the three sounds distinctly in conformity with the language community pronunciation, in order to be understood.

Finally, the learner must determine that there is a sound pattern involving nasals. In order to characterize the pattern, the learner must classify the three nasal sounds in two ways, as a single group (“nasal”) and as distinct from each other but similar to the sounds that induce the alternation (having shared “place” classifications). (The feature names are put in quotes here as a reminder that they are emergent classifications, not innate distinctive features.) Thus, acquisition of this phonological pattern involves categorization, classification, and organization.

3 Allophonic patterns

The allophonic example involves voiceless fricatives in English, the sounds /s/ and /ʃ/ as in ‘sew’, ‘show’ respectively. These are distinct, meaningful sounds with distinct articulations, as illustrated in (3). The two sounds are both considered CORONAL, and are distinguished by a second feature, [ANTERIOR]: /s/ is [ANTERIOR] while /ʃ/ is not. (For details, see Mielke *et al.* to appear; Baker *et al.* 2007; Archangeli & Baker 2008.) The acoustics of [s] and [ʃ] are dramatically distinct, with quite different frequency ranges for [s] and [ʃ], illustrated with the spectrogram in (3c). The dotted line across the middle of the figure is at 2500Hz: while [ʃ] has high amplitude frequencies starting around 1000Hz, the high amplitude frequencies for [s] are all above 3000Hz. This holds across languages, although the frequency values may vary, see Toda (2007).

(3) **Two English voiceless fricatives**

a. ANTERIOR [s]	b. CORONAL [ʃ]	c. Acoustics: “shorts”
[sip] ‘seep’	[ʃip] ‘sheep’	
[mɛs] ‘mess’	[mɛʃ] ‘mesh’	
[læsəz] ‘lasses’	[læʃəz] ‘lashes’	
		<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;">[ʃ] </div> <div style="text-align: center;">[s] </div> </div>

3.1 Fricative allophonic variation

Labov (1984); Janda *et al.* (1994); Shapiro (1995); Lawrence (2000); Janda & Joseph (2003) document a dialectal allophonic pattern. Primarily in an /stɪ/ sequence, the /s/ may be pronounced with a degree of retraction, sounding like [ʃ] as illustrated in figure (4), where the rows correspond to dialects without and with retraction, showing the perceived articulation.

(4) **Two dialects: An allophonic pattern**

<i>no retraction</i>	[st.ɪt]	‘street’	[st.ɪæp]	‘strap’
<i>retraction</i>	[ʃt.ɪt]		[ʃt.ɪæp]	

The question raised by data like these is whether the pattern is phonological, with /s/ and /ʃ/ merging in this environment, or whether it is phonetic, with /s/ and /ʃ/ remaining distinct. The difference is characterized by the two schematic rules given in (5), where (5a) shows the phonological case, with /s/ becoming [ʃ], and (5b) shows the phonetic case, with /s/ becoming [s̠], a retracted [s], distinct from [ʃ] (though perceptually quite similar). Our data show that this is a phonetic process: [ʃ] and [s̠] are distinct from each other.

(5) **Retraction rule possibilities**

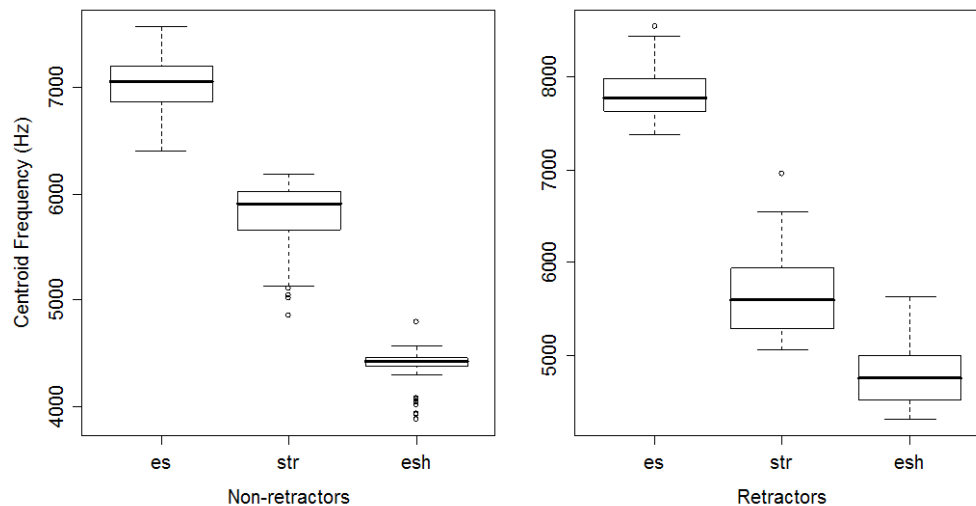
a. <i>phonological</i>	/s/ → [ʃ] / ____ tɪ	b. <i>phonetic</i>	/s/ → [s̠] / ____ tɪ
------------------------	---------------------	--------------------	----------------------

A brief summary of the study follows; see the works cited for details. Data from 32 subjects were collected, though six were discarded. Items were words beginning with /sV/, /ʃV/, /stV/, /stɪV/, and /ʃɪV/. Four tokens of each item were collected from each subject. Subjects were classified as retractors or non-retractors based on perception by trained phoneticians. Centroid frequencies of the fricatives were calculated.

Comparison of the centroid frequencies (CFs) revealed two striking facts. First, whether or not retraction is perceived, the CF of /s/ in /stɪV/ is quite a bit lower than /s/ in /stV/ or /sV/. This is shown in (6), by comparing the middle block in each graph with the two flanking blocks. (Note that the frequency scales on these two plots are not the same.) The difference between non-retractors and retractors is the distance between the CF of the retracted [s̠] and that of the /ʃ/: it is much closer together in retractors than in non-retractors.

Second, the CFs of /s/ and /ʃ/ differ depending on whether the subject retracts or not. The CF for /s/ in non-retraction environments is higher in retractors (almost 8000Hz) than in non-retractors (closer to 7000Hz), and the CF for /ʃ/ has a much broader range in retractors than in non-retractors.

(6) Acoustics of fricatives of non-retractors and retractors

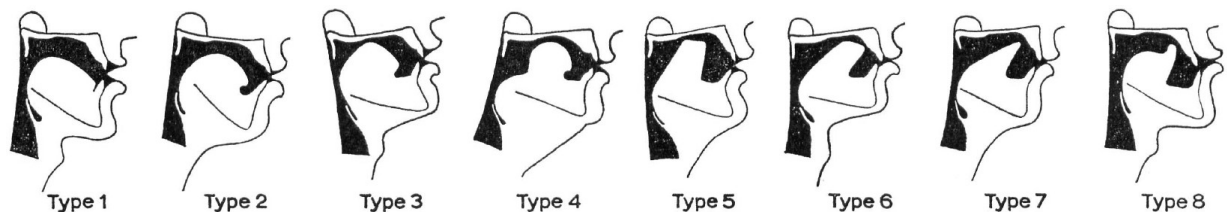


Learning includes categorizing [s] and [ʃ] as phonemically distinct in the language, as well as learning that [s̠] is phonemically the same as /s/, despite phonetic differences. The learner must figure out how to produce the different sounds, regardless of whether the differences are significant. And finally, the learner must figure out the pattern, classifying /s/ since it is targeted. Again, categorize, classify, and organize.

4 Idiophonic patterns

This brings us to our third type of case, the idiophonic patterns of /ɹ/ articulations in English; see Mielke *et al.* (to appear) for details. The first point to make is that there is no phonemic contrast with “r”s in English (unlike, e.g., Spanish which has two distinct rhotics). We might conclude, then, that there is “only one /ɹ/”. At the phonological level, this is correct. However, examination of the articulation of /ɹ/ reveals that there are multiple [ɹ]s in English. Delattre & Freeman (1968) provides the figure in (7), showing eight different ways that /ɹ/ is pronounced, depending on the speaker. Most familiar among these are the “bunched” [ɹ] with the tongue tip pointing down (Types 3-6) and “retroflexed” [ɹ], with the tip pointing up (Types 7 and 8). (Types 1,2 are the “r-less” dialects.)

(7) Only one “r”? (Delattre & Freeman 1968)



Ultrasound, audio, and video images were collected from 32 subjects pronouncing English monosyllables with pre- and post-vocalic with /ɹ/, both at word edges and in consonant clusters. (Data from 5 were rejected). Tongue tip position (up or down) made the classification of retroflexed vs. bunched, respectively. The ultrasound tracings in (8), with a superimposed palate tracing (the thin white line), face the tongue tip to the right. See Mielke *et al.* (to appear) for a full description of the study.

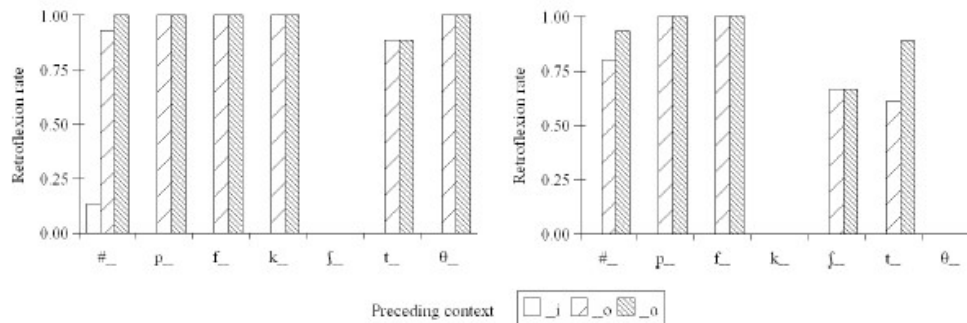
(8) **/ɹ/ articulations: subject 08**



There are three striking aspects to the data. First, many of the subjects used both articulation types, such as subject 08, (8) above. Second, in the class of bunched/retroflex speakers, the distribution of [ɹ]s occurs in stable patterns. Finally, the patterns of bunching and retroflexing are idiophonic: each subject had their own pattern.

A nice illustration of these points comes from comparing the distribution of retroflexing by subjects 08 and 17, shown in (9). First, the similarities: retroflex [ɹ] is rare before the vowel [i] and only retroflexed [ɹ] is found after [p, f, k], as long as the following vowel is not [i]. Now, the differences: following [k] and [θ], subject 08 uses only retroflex [ɹ] while subject 17 never does. By contrast, after [ʃ], subject 08 never retroflexes while subject 17 usually retroflexes.

(9) **Idiophonic /r/ patterns: Retroflexion rates for subject 08 (left) and subject 17 (right)**



These similarities and differences illustrate the idiophonic patterns of /ɹ/ articulation among the subjects in the study. No two subjects showed the same distribution of /ɹ/ articulations.

The language learner must categorize /ɹ/ as a significant sound in English along with any perceived differences, similar to the learning for [s]. Additionally, the learner has to determine how to pronounce /ɹ/. For some speakers, this means figuring out the bunched articulation; for others it means figuring out the retracted articulation. For others still, it means figuring out both articulations, and where to use each of them, and so classifying and organizing the sounds.

5 Conclusion

These examples illustrate that categorizing, classifying, and organizing sounds are critical with phonological, allophonic, and idiophonic patterns. Of interest are the different predictions made by the Emergence hypothesis and the Innateness hypothesis. Under Emergence, every

distinctive feature posited is learned based on data presented to the learner; contrasts are learned; constraints on feature behavior is learned. By contrast, the Innatist hypothesis assumes that there is an innate set of distinctive features that characterizes the phonemic sounds of a language, giving an innate set of segment contrasts and, at least in some theories, an innate set of organizational principles (e.g. feature constraints in Optimality Theory). Since these features are used for phonemic contrasts, phonetic contrasts are handled in a distinct fashion.

(10) **Hypothesis comparison**

	EMERGENCE	INNATIST
distinctive features	learned	innate
contrastive segments	learned	innate
feature-constraints	learned	innate
phonological vs. phonetic	continuum	distinct modules

Under Emergence, the mechanisms are the same regardless of the type of sound pattern being acquired. The difference lies in how the patterns function within the language. Under Innateness, acquiring sound patterns is bifurcated depending on whether the pattern can be expressed in terms of distinctive features or not. Two acquisition mechanisms are necessary, along with the different types of patterns functioning differently within the language.

(11) **Emergence vs. innatist: what must be learned**

	EMERGENCE	INNATIST			
	all sounds	[m, n, ŋ]	[s, ʃ]	[ʒ]	[ɹ, ɻ]
categorize sounds for significance	✓	✓	✓		
identify patterns	✓	✓	✓	✓	✓
classify sounds to characterize patterns	✓	—	—	✓	✓
produce differences	✓	✓	✓	✓	✓/—
map to features	—	✓	✓	—	—

Under Emergence, the same strategies are called into play regardless of the nature of the sounds and sound patterns, predicting a continuum of sound pattern types as illustrated here. Thus, universal, innate distinctive features lead to incorrect predictions and greater formal complexity, including duplication within the theory.

References

- ARCHANGELI, DIANA, & ADAM BAKER, 2008. Categorization and features: Evidence from the production and perception of American English /ɹ/. *Where do features come from?* Conference, Sorbonne, Paris.
- BAKER, ADAM, JEFF MIELKE, & DIANA ARCHANGELI, 2007. Data from English s-retraction suggest a solution to the Actuation Problem. University of Arizona and University of Ottawa.
- BLEVINS, JULIETTE. 2004. *Evolutionary Phonology*. Cambridge: Cambridge University Press.
- CHOMSKY, NOAM, & MORRIS HALLE. 1968. *The Sound Pattern of English*. Cambridge, Mass.: MIT Press.
- DELATTRE, PIERRE C., & DONALD C. FREEMAN. 1968. A Dialect Study of American r's by X-ray Motion Picture. *Linguistics* 44.29–68.

- JAKOBSON, ROMAN, GUNNAR FANT, & MORRIS HALLE. 1952. *Preliminaries to Speech Analysis*. Cambridge: MIT Press.
- JANDA, RICHARD, & BRIAN JOSEPH. 2003. Reconsidering the canons of sound change: Towards a “Big Bang” theory. In *Historical Linguistics 2001. Selected Papers from the 15th International Conference on Historical Linguistics, Melbourne, 13-17 August 2001*, ed. by Barry Blake & Kate Burridge, 205–219. Amsterdam: John Benjamins Publishing.
- JANDA, RICHARD D., BRIAN D. JOSEPH, & NEIL JACOBS. 1994. Systematic Hyperforeignisms as Maximally External Evidence for Linguistic Rules. In *The Reality of Linguistic Rules*, ed. by S. Lima, R. Corrigan, & G. Iverson, 67–92. Amsterdam: John Benjamins.
- LABOV, WILLIAM. 1984. Field methods of the project on language change and variation. In *Language in Use*, ed. by John Baugh & Joel Scherzer, 28–53. Englewood Cliffs, NJ: Prentice Hall.
- LAWRENCE, WAYNE. 2000. /str/ → /ftr/: Assimilation at a distance? *American Speech* 75.82–87.
- MCCARTHY, JOHN, & ALAN PRINCE. 1993. *Prosodic Morphology I: Constraint Interaction and Satisfaction*. Technical Report #3. Rutgers University: Rutgers University Center for Cognitive Science.
- MIELKE, JEFF. 2004. *The Emergence of Distinctive Features*. The Ohio State University dissertation.
- . 2005. Ambivalence and ambiguity in laterals and nasals. *Phonology* 22.2.169–203.
- , ADAM BAKER, & DIANA ARCHANGELI. to appear. Variability and homogeneity in American English /ɹ/ allophony and /s/ retraction. In *Variation, Detail, and Representation: LabPhon 10*. Berlin: Mouton de Gruyter.
- MOHANAN, K. P., DIANA ARCHANGELI, & DOUG PULLEYBLANK. to appear. The emergence of Optimality Theory. In *Reality Exploration and Discovery: Pattern Interaction in Language and Life*, ed. by Linda Uyechi & Lian-Hee Wee. Stanford University: Center for the Study of Language and Information.
- PORT, ROBERT. 2007. How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology* 25.143–170.
- PRINCE, ALAN, & PAUL SMOLENSKY. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical Report RuCCS-TR-2. New Brunswick, NJ: Rutgers University Center for Cognitive Science.
- SHAPIRO, MICHAEL. 1995. A case of distant assimilation: /str/ → /ftr/. *American Speech* 70.101–107.
- TODA, MARTINE. 2007. Speaker normalization of fricative noise: Considerations on language-specific contrast. *International Congress of Phonetic Sciences XVI*.825–828.
- TRUBETZKOY, NICOLAI SERGEYEV. 1936. Essai d’une théorie des oppositions phonologiques. *Journal de psychologie normale et pathologique* 33.5–18.

Diana Archangeli
dba@u.arizona.edu

Development of communication in a social/emotional context

Arlene S. Walker-Andrews
The University of Montana

By seven months, infants are adept perceivers of others' emotional expressions, even when the "other" is unfamiliar. Add contextual support - ongoing dialogue, a familiar person, multimodal and dynamic expressions - and three-month-olds demonstrate that they can discriminate, generalize, and respond differentially to expressions. Study of infants' perception of emotional expressions provides a productive vantage point from which to study infants' developing abilities to detect structure and meaning in the environment. Studies of infants' perception of speech suggest that comprehension develops from a dynamic, complementary relationship between the infant and environment. As with emotion, infants detect invariant, amodal or redundant information in bimodal speech. The interchange between infant and others is characterized by dynamic and reciprocal interplay between intermodal perception, selective attention and learning, coupled with specific structure marking intentional communication. Speech acts, like expressive signals, comprise rich, multimodal events that infants mine as they respond to the perceived affordances for behavior.

What can one say about infants' perception of emotional expressions and how research on this topic contributes to an understanding of the development of language? Taking a cue from ethology, I subscribe to the view that emotional expressions are a form of communication, like language, that functions as a guide to action. That is, expressions are not only "signs of emotion", but social signals that afford information about another individual's likely behavior (his or her intentions) and, therefore, point the way for one's own actions. Facial expressions are only a part of a set of behaviors (including vocal expressions, gestures, touch, and dynamic properties such as contingency) that convey the intentions of the actor. Given this as a starting point, I suggest that infants will be most sensitive to the emotional expressions provided in dynamic, multimodal contexts, in familiar situations, and by significant others with whom the infants has established patterns of interaction. Intentionality is likely to be detected first in intense and frequent interaction such as those provided by the parents of infants.

Selective attention

Nearly 30 years ago, Bahrick, Walker, and Neisser (1981) conducted a study focused on infants' selective attention, and the importance of intermodal correspondences to perception and attention. We live in, and in fact are part of, a multimodal world. In the typical environment, several objects and events may be perceptible at the same time and in the same general direction. We do not "experience any particular difficulty in such situations: perceptual selection is smooth and easy" (Bahrick et al, 1981, p. 377). Likewise, the young infant encounters a world of objects, events, people, and places that are specified multimodally. Even the self is specified across modalities: we can see our own arms and legs, hear our own voices and growling stomachs, taste salty tears, experience feedback when we touch one finger to another. Historically, researchers proposed that perceivers must learn to pair and integrate sense-specific information in order to interpret it as meaningful. The typical question was, how do perceivers "bind together" these unrelated bits of information. Commonly the mechanism of "association" was invoked to explain how the perceiver comes to integrate information to yield a unified percept.

Conversely, think about the redundancy provided to an observer when an object is encountered, such as when an object strikes a surface, thereby producing visual and acoustic information united by synchrony and co-location, as well as correspondences between the type of sound made by a particular substance. This redundancy allows "optimal deployment of attention and the discovery of higher order perceptual structure" (Bahrick & Lickliter, 2002, p. 156). Bahrick et al (1981) examined infants' selective attention to complex, multimodal events, in which infants detected such redundancy. In essence, the four-

month-old infants were provided with an event seemingly impossible to follow visually. We superimposed films of two events (a hand-clapping sequence; and the noisy movements of a yellow plastic slinky). That is, we overlapped these filmed images on a backprojection screen in front of the infant and played the soundtrack to one of the events. This is a very chaotic-looking event: two different actions and sets of objects moving in the same space. The question was whether infants could use the relationship of the soundtrack to disambiguate and follow one of the events. An analogy might be drawn to a common experience when one is standing at a window: you can focus on the windowpane itself, staring at your own reflection, or you can look through the window to the scene beyond. Infants viewed this strange combination, and every 20 seconds, the projectors (which were placed on lazy Susans) were rotated by a few degrees, so that the two events were seen projected side-by-side. We monitored infants' looking to the two events when they were separated spatially to determine whether infants were able to follow one of them when they were superimposed by using the soundtrack. Did the accompanying sound help the infant focus on the specified event?

The results were impressive. We found that infants looked more often (about 67% of the time) at the event that had been silent during the superimposition phase. With a number of control experiments, we satisfied ourselves that infants were perceiving a single, unified event during the superimposition. That is, in one experiment, we presented infants a single event by blocking off one of the projectors, and then showed two events side-by-side on the tests. As before, infants looked about two-thirds of the time to the new event. In another experiment, we showed the infants superimposed events as before, but defocused them sufficiently so that only blurred color and motion without form were evident. Infants showed no looking preferences when the events (now in focus) were presented side by side.

This set of experiments illustrates infants' ability to use redundant information to disambiguate events, as the infant must do in real-world situations. When the infant is lying in the crib and Grandma and Grandpa enter the room, swooping down close, exclaiming excitably, and stroking his or her face, the infant is able to hone in on which person is talking by attending to a myriad of correspondences between speaker and voice. The infant does not hear a voice and see a face (or hear two voices and see two faces) and somehow glue these together, rather the infant experiences a person interacting with him/her and responds accordingly.

As shown in the Bahrack et al (1981) experiments, the infant benefits from the amodal, invariant relationships that serve to separate events. Common temporal patterns, rate of action, temporal synchrony relations, and even the appropriateness of the type of sound for the visual information is sufficient for infants as young as 4 months to selectively attend to a particular event. This ability allows infants to explore the environment and attend to relevant features and actions in the world. The developmental task is to differentiate increasingly more specific information by detecting invariant patterns such as those present in complex, multimodal events.

Perception of emotion

Do the data on infants' perception of emotion bear out expectations that infants will be most sensitive to the emotional expressions provided in dynamic, multimodal contexts, in familiar situations, and by significant others? It appears that, by seven months, infants are adept perceivers of others' emotional expressions. A number of researchers have examined infants' abilities to discriminate and generalize emotional expressions. For example, given a series of facial expressions drawn from the same discrete category of emotion (say, happy), infants visually dishabituate when they are confronted with a new expression (sad), but they fail to show increases in looking time to yet another happy expression, even when it is presented by a different person altogether. Infants of this age also show intermodal matching of facial and vocal expressions. When they view two filmed facial expressions, side-by-side, along with a single soundtrack matching one of the expressions, infants will look proportionately longer at the sound-specified facial expression. By seven months, infants detect the information that is invariant across several different exemplars of one expression, and they detect the intermodal correspondences that characterize a single facial-vocal expression. Infants don't simply use synchrony information in this task, as presenting the faces upside-down (Walker, 1982) or obscuring the mouth so lip-voice synchrony is hidden (Walker-Andrews, 1986) does not disrupt the intermodal matching shown by seven-month-old

infants. Even when much of the featural information is removed (by using point-light displays), infants make matches based on dynamic intermodal correspondences (Soken & Pick, 1992).

These are remarkable achievements, but it is even more interesting to contemplate the abilities of younger infants when contextual information is provided. For example, Montague and Walker-Andrews (2001) investigated infants' discrimination of expressions when these were embedded in a familiar game of peekaboo, selected because peekaboo is a familiar game with specific expectations. In fact, it is a game of affect. We found that four-month-olds could discriminate happy, sad, angry, and fearful expressions in this familiar game, and they responded to the expressions in differentiated ways. In the experiment we conducted, infants viewed an unfamiliar woman act out the typical peekaboo game, but on the fourth (and eighth) reappearance from behind her hands, the woman presented an unexpected expression for some infants. That is, one-quarter of the infants viewed a sad expression on the fourth and eighth trials, another quarter saw angry, another group saw fearful expressions, and one-quarter continued to observe the typical happy/surprise peekaboo. As expected, infants who viewed the typical happy expression gradually decreased looking time across the set of eight trials, much like in a habituation experiment. Those infants who saw sad decreased their overall looking time to an even greater degree. Infants who saw fearful expressions on these target trials, increased their looking time on the first incidence, but decreased looking on the second fearful presentation. Infants who saw an angry expression on the fourth trial increased their looking time, and remained at higher levels of looking throughout the rest of the game. It was as if the angry expression led to "vigilant" behavior on the part of the infants. These results emphasize the importance of contextual information to the recognition of emotional expressions.

Infants also show earlier discrimination, generalization, and intermodal matching when expressions are portrayed by familiar people. Many studies on mother-infant interaction provide evidence that infants are sensitive and active participants in these interactions. For example, Haviland and colleagues (Haviland & Lelwica, 1987) examined the responses of 10-week-old infants to their own mothers' live presentations of happy, sad, and angry facial/vocal expressions. These infants responded differentially and contingently to maternal emotional signals. Haviland also documented early emotional responsiveness among mother-infant dyads, finding that mother exploit this mode of interaction as a context for socialization of infants' expressiveness to bring them into line with cultural expectations (Malatesta & Haviland, 1985).

We (Kahana-Kalman & Walker-Andrews, 2001) investigated infants' intermodal matching when familiar persons depict the expressions. Infants were presented simultaneously with two filmed facial expressions (happy and sad) accompanied by a single vocal expression that matched one of the two facial expressions. Infants as young as three months showed intermodal matching when they observed vocal and facial expressions portrayed by their own mothers versus those posed by an unfamiliar woman. They looked longer to the sound-specified facial expressions of their mothers (whether presented in or out of synchrony). In addition, infants were rated as experiencing more positive affect and as more interested and engaged, particularly when the emotion displays were portrayed by their own mothers. Infants who observed their own mothers are rated as more positive and more engaged when happy was the sound-specified emotion, and they spent more time smiling as well. The types of smiles (full, bright, faint) also differed as expected across groups and conditions (familiar, unfamiliar, asynchronous, happy and sad).

Montague and Walker-Andrews (2002) completed a subsequent study in which infants were presented happy, angry, and sad facial and vocal expressions depicted by their mothers, fathers, and unfamiliar males and females in an intermodal preference procedure. In this case, infants showed intermodal preferences for mothers' expressions, but for fathers the results were more mixed. Overall, they did not show intermodal preferences for the paternal expressions, but a closer look yielded an interesting result. Those infants whose fathers were more involved in caregiving activities (as indicated by a time-diary interview and child-care activity survey) showed the same intermodal preferences shown for the maternal expressions. These findings indicate that early in their development, infants are sensitive to contextual information that may facilitate detection of the meaning of others' emotional expressions.

Finally, infants as young as three months generalize facial and vocal expressions when these are presented by familiar persons. We (Walker-Andrews, Mayhew, Coffield, & Krogh-Jespersion, submitted) habituated infants to films of their parents alternately presenting happy or sad facial-vocal expressions. That is, infants viewed their mother and then their father acting out an expression during a visual habituation sequence. When the infants decreased their looking time, they were presented either (a) their

mother continuing to act out the habituated expression, (b) their mother acting out a new expression, (c) a stranger acting out the familiar expression, or (d) a stranger acting out a new expression.

In this study, infants demonstrated that they discriminated and generalized the parental expressions. First, it is interesting that infants took more than six minutes to habituate to the parental expressions, an exceedingly long time for infants to visually explore such a sequence. In addition, infants showed different patterns of dishabituation, depending on the expression and the identity of the actor. In brief, infants who viewed parental happy expressions during the habituation sequence failed to increase their looking time when their mother modeled the happy expression on the posttest; showed an increase when their mother modeled a sad expression; showed an increase when the female stranger modeled the happy expression; and showed an increase in looking time when a stranger modeled a sad expression (all significant increases). Infants who viewed sad expressions during the habituation sequence also increased their looking time to changes in person and/or expression. Infants who continued to view their mother depicting a sad expression did not show a significant increase in looking time. Those who viewed their mother depicting a novel happy expression increased their looking time by approximately a minute; those who viewed a stranger depicting the familiarized sad expression increased looking time by approximately 90 seconds; and those who viewed a stranger depicting a novel happy expression increased looking time by approximately a minute.

A second experiment conducted with a new group of infants of the same age, using the same films yielded a different pattern of results. Because these infants were habituated to another infant's parents acting out happy or sad, all expressions were depicted by unfamiliar adults. Infants discriminated changes only in the cases where both person and expression changed at test. As a final check, we enrolled a group of seven-month-olds in the study. These infants discriminated and generalized the emotional expressions as expected, even though the emotional expressions were depicted by unfamiliar persons.

Results such as these show that infants are most sensitive to the emotional expressions provided in dynamic, multimodal contexts, in familiar situations, and by significant others with whom the infant has established patterns of interaction. Study of infants' perception of emotional expressions provides a productive vantage point from which to study infants' developing abilities to detect structure and meaning in the environment. I would argue that familiarity, for example, serves as an intrinsic contextual factor that adds meaning to the task of early recognition of emotion. It is not just that parental expressions are more common in an infant's environment, but they also may be especially informative with respect to ensuing actions. Infants may be more motivated to attend to the affective behaviors of their parents, because these may foreshadow more specific outcomes for them. That is, typically, a caregiver's smiles are likely to be followed by positive interactions or experiences of being held; in contrast, negative expressions may be frequently followed by episodes in which the infant is ignored. Thus, it is possible that infants more readily recognize the affective signals of mothers (and fathers) because these entail more idiosyncratic patterns of ensuing interactions. Emotion expressions are socially communicative, and infants are encountering these social signals early on, from their very early exposure to mother's approving smile and disapproving frown. Each signal has its own affordance: a happy expression signals "readiness for friendly interaction" (Izard, 1993, p. 634), an angry expression signals something entirely different. Infants learn to detect the important social information, the putative intentions of others, very early on.

Parents, too, respond sensitively and contingently to the infant's active overtures, creating a dynamic and self-organizing communication system. Young infants experience an organization in the social world via early imitative exchanges and via the dialogue of "vitality" affects that the partners share. The topic of conversation in social exchanges is affect, and interactions focus on sharing and regulation of affect and excitement. I have argued before that this also contributes to an infant's developing sense of self, as well as of the "other". As infants are learning about others' expressions they are experiencing their own as well, gaining additional information for continuity of self and an appreciation for others. As Stern (1985) puts it, there comes to be "affect attunement". Neither mother nor infant imitates the other's expression exactly, but they each convey the same quality of feeling. It is perhaps analogous to verbal interactions in which a parent expands or recasts an infant's two-word utterance, duplicating the underlying meaning but altering the form of the utterance. "Affect attunement is an ongoing, multimodal interaction that highlights self and other and the shared perception of affective affordance" (Walker-Andrews, 1992, p. 132).

Word learning

Lakshmi Gogate and I referred to this sort of entrainment when we described word learning during infancy (Gogate, Walker-Andrews, & Bahrick, 2001). Early lexical development can be described as a process of continuing reciprocal interactions between the organism and the environment, in this case infant and mother (or other communicating individual). In fact, much of vocal development can be viewed in this light. For example, Gros-Louis, West, Goldstein, and King (2006) found that mothers of eight-month-olds naturally provide not only contingent responses to infants' vocalizations, but also responses that are specific to particular vocal types, which could scaffold vocal development. The mothers gave specific kinds of verbal feedback to vowel-like and consonant-vowel clusters, which differ not only acoustically, but also are produced by infants at different developmental stages of vocal production. Mothers responded differentially with interactive-play vocalizations, depending on the type of vocalizations made by the infants. These responses, in addition to providing contingency, "provide some information to the infant about the "effectance" of infants' vocal production. Through differentiated maternal responding, mothers encourage the use of particular sounds, giving them meaning and framing the interaction" (p. 514).

In Gogate's work, infants' abilities to learn the arbitrary relations between speech sound and objects is examined. For example, Gogate and Bahrick (1998) found that seven-month-olds learn the relation between speech sounds /a/ and /i/ and moving objects, when the timing of the vocalization coincides with that of the object. But it is not a simple "association", as infants fail to learn these relations when temporal contiguity between vowel sound and static object is presented. Eight-month-olds detect arbitrary relations between more complex monosyllables /tan/ and /gah/ and moving objects when temporary synchrony, the "bootstrap", is present.

Moreover, parents provide a variety of multimodal naming contexts during everyday interactions with their infants. Mothers use gestures to highlight meaning by pointing to or touching an object when referring to it (Zukow-Goldring, 1997), they acoustically highlight specific words through prosody, use shorter sentences, and place a word in sentence final position. They provide redundant information such as temporal synchrony when they simultaneously name and show objects to infants (especially between five and eight months), while they rely on simply pointing to objects for older infants (nine to 17 months), and just name and hold objects for much older infants (21 to 30 months). The suggestion is that mothers regulate the naming context by structuring the environment according to the infant's level of perceptual-lexical competence.

Summary

In conclusion, the combination of infants' sensitivity to intermodal correspondences, enhanced by contextual information, and the multimodal character of emotional expressions leads to early recognition of those expressions. The early emergence of emotion recognition may provide the foundation for many other developing competencies, including the ability to predict and respond appropriately while engaged in social interactions.

References

- Bahrick, L. E., & Lickliter, R. (2002). Intersensory redundancy guides early perceptual and cognitive development. *Advances in Child Development and Behavior*, 30, 153-187.
- Bahrick, L. E., Walker, A. S., & Neisser, U. (1981) Selective looking by infants. *Cognitive Psychology*, 13, 377-390.
- Gogate, L. J., & Bahrick, L. E. (1998) Intersensory redundancy facilitates learning of arbitrary relations between vowel-sounds and objects in 7-month-olds. *Journal of Experimental Child Psychology*, 69, 133-149.
- Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001) The intersensory origins of word comprehension: An ecological-dynamic systems view. *Developmental Science*, 4, 1-37.

- Gros-Louis, J. G., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, 30, 509-516.
- Haviland, J. M. & Lelwica, M. (1987). The induced affect response: 10-week-old infants' responses to three emotion expressions. *Developmental Psychology*, 23, 97-104.
- Izard, C. E. (1979). *The maximally discriminative facial movement coding system (MAX)*. Newark: University of Delaware, Information Technologies and University Media Services.
- Kahana-Kalman, R., & Walker-Andrews, A. S. (2001). The role of person familiarity in young infants' perception of emotional expressions. *Child Development*, 72, 352-369.
- Malatesta, C. Z. & Haviland, J. M. (1985). Signals, symbols, and socialization: The modification of emotional expression in human development. In M. Lewis & C. Saarni (Eds.), *The socialization of emotions* (pp. 89-116). New York: Plenum.
- Montague, D. P. F., & Walker-Andrews, A. S. (2001) Peekaboo: A new look at infants' perception of emotion expressions. *Developmental Psychology*, 37, 826-838.
- Montague, D. P. F., & Walker-Andrews, A. S. (2002). Mothers, fathers, and infants: The role of person familiarity and parental involvement in infants' perception of emotion expressions. *Child Development*, 73, 1339-1352.
- Soken, N. H. & Pick, A. D. (1992). Intermodal perception of happy and angry expressive behaviors by seven-month-old infants. *Child Development*, 63, 787-793.
- Stern, D. N. (1985). *The interpersonal world of the infant*. New York: Basic Books.
- Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology*, 33, 514-535.
- Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? *Developmental Psychology*, 22, 373-377.
- Walker-Andrews, A. S. (1992). A developing sense of self. *Psychological Inquiry*, 3, 131-133.
- Walker-Andrews, A. S., Mayhew, E., Coffield, C. & Krogh-Jespersen, S. (submitted). Young infants' generalization of emotional expressions: Effects of familiarity.
- Zukow-Goldring, P. (1997) A social ecological realist approach to the emergence of the lexicon: educating attention to amodal invariants in gesture and speech. In C. Dent-Read & P. Zukow-Goldring (Eds.), *Evolving explanations of development: Ecological approaches to organism-environment systems* (pp. 199-252). Washington, D.C.: American Psychological Association.

Face and Speech: How and when do infants understand Ethnicity?

Olivier Pascalis and Lesley Uttley

LPNC, Grenoble, France and Department of Psychology, University of Sheffield

1. What is ethnicity?

An ethnic group is a group of human beings whose members identify with each other, through a common heritage: using cultural, linguistic, religious, behavioural, or biological traits as indicators of contrast to other groups. As adults we often use faces as a cue for ethnicity, even if we know that it is not necessarily accurate. However, faces might be our only means of studying perception of ethnicity during infancy.

Faces are one of the most predominant visual stimuli in children's environments. From birth onwards, children encounter hundreds of faces. These faces vary in terms of not only identity, but also gender, age, attractiveness, species, and race. Given the adaptive significance of the face processing ability, the hypothesis about an innate disposition to such an ability is appealing. However, evidence has accumulated in the last several decades that suggests the prominent role of experience in shaping children's face processing expertise, which in turn forms the foundation for later face expertise in adulthood.

2. Face preference

It has been established that newborn human infants demonstrate a visual preference for both real and schematic human faces over almost any other category of stimulus (Fantz, 1964; Goren, et al., 1975; Johnson, et al., 1991; Valenza, et al., 1996). Whilst this may not provide conclusive evidence for an innate face processing module, it is suggestive of a specialized cognitive system operating from very early in life.

3. Categorization

Categorization is 'a mental process that underlies one's ability to respond equivalently to a set of discriminably different entities as instances of the same class (Quinn & Slater, 2003). It is the construction of a prototypical average of a class of stimuli. Face processing develops quickly from birth with infants receiving more and more experience of only a handful of face exemplars, mainly their parents. Recently, de Haan et al. (2001) demonstrated that infants begin to show evidence of face prototype formation at 3 months of age. Before this age, face recognition seems to be exemplar based: each individual face is separately encoded. Quinn et al. (2002) hypothesized that the face representation of 3-month-olds may be biased toward female faces as their primary caregiver is usually the mother. They found support for their hypothesis by demonstrating that 3-month-olds infants raised by their mother prefer to look at female faces when paired with male faces. Quinn et al. also identified and tested a small population of 3-month-olds who had been raised by their fathers. In this instance, a preference for male faces was observed. Kelly et al. (2005) further highlighted the role of experience in early infancy, demonstrating that 3-month-old Caucasian infants prefer to look at faces from their own racial group when paired with faces from other racial groups in a visual preference task. Caucasian newborns tested in an identical manner demonstrated no preference for faces from either their own- or other-racial groups. They concluded that the preference observed in 3-month-old infants is a direct consequence of predominant exposure to faces from their own race early in infancy. Two more recent studies have shown that this early preference is not confined to Caucasian infants. Kelly et al. (2007) replicated their findings with 3-month-old Chinese infants tested in China, and Bar-Haim et al. (2006) found a preference for own-race faces in 3-month-old Israeli and Ethiopian infants, also tested in their native countries. Bar-Haim et

al. also tested a population of infants of Israeli-born, Ethiopian infants who had received exposure to both African and Caucasian faces. These infants showed no preference for faces from either racial-group. Collectively, these results clearly demonstrate how the early face processing system is influenced by the faces observed within the infants' visual environment.

4. Face recognition:

Face recognition entails recognizing that a specific face has been encountered before and assessing its familiarity. The ability to learn and recognize individual faces is paramount for the development of attachments and social interaction. Recognition of the mother is particularly important for the development of attachment and emotional bonds between mother and child (Bowlby, 1969). Psychologists have attempted to determine when and how the mother's face is first recognized by the infant using both the real face and pictures of the mother's face. It has been demonstrated that mother's face can be recognized from 3 days of age (Field, Cohen, Garcia and Greenberg, 1984; Bushnell et al., 1989; Pascalis et al., 1995). This representation is not unimodal. A recent study by FatmaSaï (2005) elegantly demonstrated that newborns only recognise their mother's face if a postnatal exposure to the mother's voice-face combination was available. It appears that the mother's face is in fact learned in conjunction with the mother's voice, which has been heard during gestation. If the infant is denied the auditory input of the mother's voice after birth, recognition of the mother's face is not demonstrated at this stage.

The mother's face is, however, a special case as it is positively reinforced and is interrelated with attachment. Only a handful of studies have investigated recognition of strangers' faces in newborns. It has been shown that 4-day-old infants habituated to a photograph of a stranger's face can recognize it immediately and after a retention interval of 2 minutes (Pascalis and de Schonen, 2004; Turati, Cassia, Simion, and Leo, 2006). Neonates are able to learn and recognize another face as early as 4 days of age even if the face is not presented with the full multimodal aspect. During the first 6 months of life face processing then improves rapidly.

Faces convey not only information about the identity of a person but also a variety of social information such as emotional expression, facial speech or gaze direction: faces are multi-dimensional visual stimuli. Face perception starts to develop very early and faces provide an early channel of communication between infant and caretaker prior to the onset of language. They provide rich sources of visual information with social significance. Investigating how face processing develops in multi-racial environment in conjunction with social judgments of others provides a unique opportunity to study the influence of perceptual, cognitive and social information on the development of both recognition and social attitude to others.

During the Napoleonic Wars, a French ship was wrecked off the Hartlepool coast. The Hartlepool fishermen were concerned about the possibility of French infiltrators and feared an invasion. Among the wreckage lay one wet and sorrowful looking survivor, the ship's pet monkey dressed in a military style uniform. Unfamiliar with what a Frenchman looked like, the fishermen came to the conclusion that this monkey was a French spy and should be sentenced to death. The unfortunate creature was to die by hanging.

This story represents one extreme example of poor discrimination abilities and of negative attitudes toward other ethnic groups. The 'Other-Race Effect' (ORE) describes the phenomenon that humans find it easier to discriminate between faces from their own ethnic group than between those from other ethnic groups. In a recent meta-analysis, Meissner and Brigham (2001) found that people were 1.4 times more accurate in identifying previously seen own-ethnic group faces compared with previously seen other-ethnic group faces. The ORE is assumed to be a consequence of experience with faces from unique ethnic groups in our environment and has important theoretical and applied implications. From a theoretical perspective, the ORE reflects how humans process individual faces and group them in different social categories (e.g., categories based on ethnicity, gender, and age) and provides unique insights into the mechanisms and processes that underpin human face processing in particular and social interaction more generally.

The ontogeny of phonemic perception is characterized by a decline in the discrimination of speech sounds not present in one's native language. Thus, native language experience functions to tune,

maintain, and facilitate the perception of phonemes. Currently, there is substantial evidence suggesting that this specialization occurs between 6 and 12 months of age and is critically dependent on perceptual exposure to native, relative to non-native phonemic contrasts (Werker and Tees, 2005). In addition to experience-driven maintenance of native contrasts, there is also evidence of facilitation of these native contrasts (Kuhl, et al., 2006). Nelson (2001) hypothesized that similar to the development of phonemic perception, face processing abilities develop based on the types of faces present in the visual environment. It then tunes toward the predominant faces in the environment. According to this account, the infant begins life with a crude and unspecified face representation which is then subject to modification as a result of the category (i.e. human) of facial input received. This notion is best understood within the framework of the multidimensional face space model described by Valentine (1991). Valentine proposes a norm-based coding model in which faces are encoded as vectors according to their deviation from a prototypical average. As argued by Nelson, at birth the dimensions of the prototype are considered to be broad and largely unspecified with ensuing development of the prototype dependent upon facial input. The resulting dimensions will differ according to the input received with certain salient, individuating dimensions carrying more “weight” than others. Predominant exposure to faces of a specific species, gender, or race early in life will cause the dimensions of one’s prototype to become “tuned” towards such faces; it is now named “perceptual narrowing.”

In order to test whether experience tunes face processing, we investigated the ability of 6- and 9-month-old infants to recognize faces from their own (human) and other species (Rhesus Macaque) using a standard infant recognition paradigm. Infants at both ages were able to demonstrate recognition with human faces, looking longer at a new human face compared to a previously seen human face. However, when tested with the monkey faces, only the 6-month-old group showed evidence of recognition. That is, 9-month-olds were unable to identify which monkey face they had seen before. These findings suggest that the face system becomes ‘tuned’ to human faces between 6- and 9-months of age (Pascalis, de Haan, and Nelson, 2002). Recent developmental studies of the ORE confirm that the prototype of the human face is “tuned” during early infancy in response to differential face input. We have shown that while 3- and 6 month-old Caucasian infants are able to discriminate between pairs of Caucasian or Chinese faces, 9-month-old Caucasian infants show difficulties in discriminating between Chinese faces (Kelly et al., 2007). The convergence of findings from our studies on own-species (Pascalis et al., 2002) and the ORE (Kelly et al., 2007) indicate that 6- to 9-months of age represents an important time of transition in the face processing system. If a certain type of face (other species or other races) is not experienced prior to this period then we appear to lose our ability to discriminate between individual faces within those groups.

Infants are also sensitive to events which are bimodally specified. This is the integration of information from two sense modalities into a single percept. Using such information, young infants are able to correctly match sound and vision to identify the appropriate moving object (Spelke, 1979), the gender of the speaker (Poulin-Dubois, et al., 1995; Patterson and Werker, 2002), as well as the age of the speaker, and also to discriminate between emotions (Walker-Andrews and Lennon, 1991). They are also able to make assumptions about categories by matching intermodal information from pictures and sounds that they have little or no experience with early in life. To test whether perceptual narrowing operates in the development of intersensory perception, Lewkowicz and Ghazanfar (2006) investigated infants’ ability to match face and voice pairings across development. Infants aged 4-, 6-, 8-, and 10-months viewed two side-by-side images of vocalizing monkey faces while listening to one of the two vocalizations. Results reveal that only the younger infants were able to correctly match the vocalization they heard with the monkey face making that vocalization, as indicated by a looking preference for the sound-face match. However, by 8-10 months of age, no sound-face match was made. Evidence in support of the multi-modal nature of perceptual narrowing also comes from research investigating the development of language discrimination using silently presented articulations (Weikum et al., 2008). Weikum and colleagues report that English and French monolingual 4- and 6-month-old infants are able to discriminate both French and English silent articulations. However, after 8-months of age, only French-English bilingual infants discriminate these same silent articulations. Combined, these investigations provide direct evidence that perceptual narrowing operates in the development of intersensory perception in addition to previous unimodal visual and auditory perception.

So far, we can conclude that infants have a representation and an expectation of other humans (Bonatti et al., 2002) that changes rapidly with experience during the first years of life. How precise is this human representation? Does it extend to language and culture? Ethnicity and language are examples of naturally occurring categories, with races differing in face morphology, skin tone, and speech. Will infants expect a face of their own race to speak their native language and a face of another race to speak a non-native language?

5. Initial study

In our first study, we were interested in finding out if we could reverse the preference for own-race faces observed by Kelly et al. (2005) by adding other information such as voices.

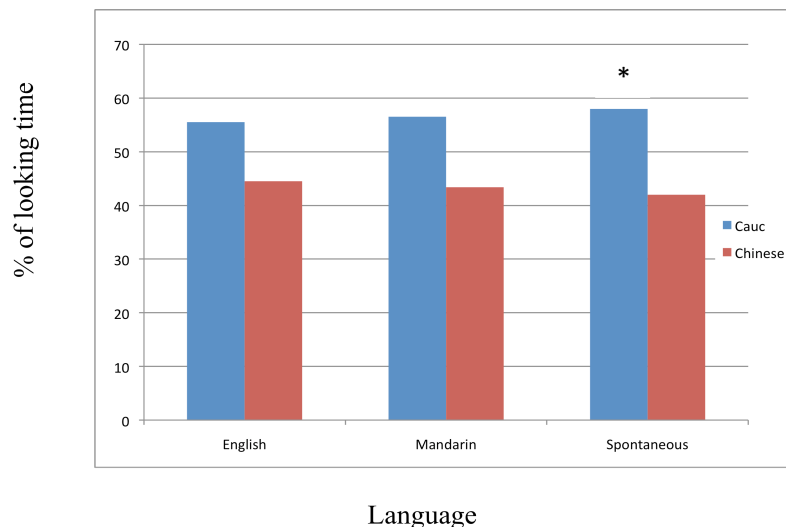
3- and 6-month-old Caucasian infants from English speaking families were first presented with a voice speaking either English or Mandarin, while seated in front of a blank screen. After 10-seconds of sound exposure, a pair of faces was displayed until 10 seconds of fixation time had elapsed. The pair was a Caucasian female and a Chinese female (see figure). Each child saw four pair of faces, 2 preceded by English and 2 by Mandarin.



Examples of stimuli used.

6. Results

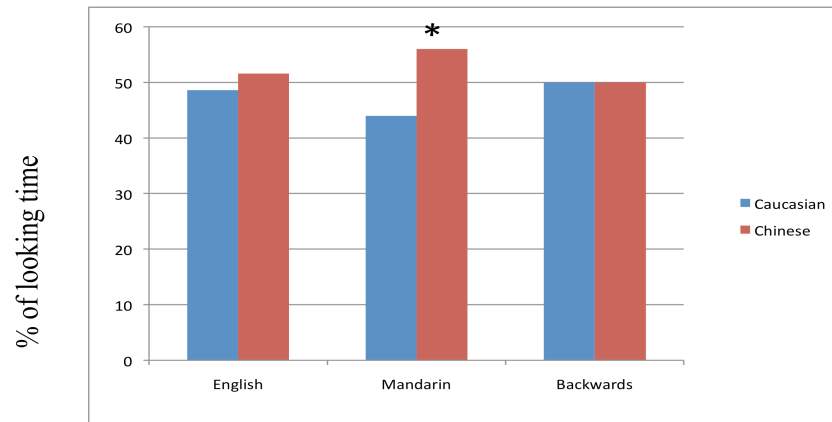
At 3 months of age, we observed a preference toward Caucasian faces even if the preceding language was Mandarin. However, it was significant only when no language was associated.



3-month-olds looking time toward Chinese and Caucasian faces when hearing English or Mandarin. The spontaneous graph is from Kelly et al., (2005).

It seems that the visual attraction for familiar faces is very strong at 3 months of age and is mildly influenced by the language.

At 6 months of age, a null preference is observed with the familiar language or with speech played backwards, but a significant preference toward Chinese faces is observed when Mandarin has been heard before the presentation.



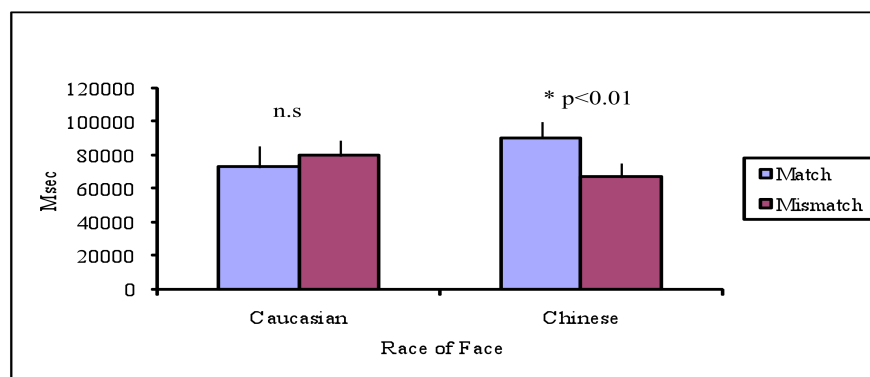
Language

6-month-olds looking time toward Chinese and Caucasian faces when hearing English or Mandarin or backwards speech.

The spontaneous preference observed for Caucasian faces at 3 months of age has disappeared; 6-month-olds are interested in both the familiar face and the novel face, but when a novel language has been presented first, their visual attention is directed toward the new face. These results suggest that by 6 months of age infants have a representation of their own race that may involve language. However, the null result observed when English was played might be due to interference between visual preference and sound.

We decided to investigate this issue using an infant control task in which infants were presented with 4 faces, one at a time until the child looked away for 2 seconds or more. During the visual presentation, a female voice was speaking either in English or in Mandarin. It was a between-participant design as one group of infants was presented with Caucasian faces and a second group with Chinese faces.

We tested 3-, 6- and 9-month-olds. All three age groups presented the same pattern of results illustrated below.



Looking time towards Chinese and Caucasian faces with language match or mismatch

3-, 6- and 9-month-old infants spent the same amount of time looking towards an own-race face speaking either a familiar or an unfamiliar language, suggesting that language did not influence their visual attention toward those faces. They are however more interested by an other-race face speaking a new language than a familiar language. It can be interpreted as a double novelty effect: a new face-type and a new language equals increased looking time.

7. Discussion

During the first few months of life, infants see few faces, usually from the same race and speaking only one language. This experience is, however, enough to learn something about the face-speech relationship. In recent years evidence has emerged that the face processing system undergoes a similar process of perceptual narrowing and tuning over the first year of life similar to that of language. Our results provide further evidence for parallels in the development of those two cognitive systems. Why do discriminatory abilities in the visual and auditory systems appear follow a common developmental trajectory? A recent study suggests that, in an immediate paired comparison task, newborns discriminate static images of their mother's face from a stranger's face if postnatal exposure to the mother's voice-face combination was available. If the infant is tested prior to seeing their mother's voice-face paired, discrimination of the mother's face is not demonstrated (Sai, 2005). This, combined with research finding perceptual narrowing when infants view silent articulations (Weikum, et al., in 2008), suggest that face and speech perception are linked from an early age. We observe a developmental specialization in perceptual discriminatory abilities during the first year of life. This specialization corresponds to improved discriminatory efficiency for stimuli predominant in the child's environment compared to a decline in discriminating stimuli not present in the surrounding environment. However, what mechanism or mechanisms are responsible for this narrowing and/or maintenance/facilitation of discrimination with experience is not well understood.

Could our results be the basis of the representation of ethnicity? It seems that a face-speech representation of humans emerges during the first year of life. But face processing is still very flexible during childhood as illustrated by the Sangrigoli et al. (2005) study that shows that Korean children adopted into Caucasian families in a predominantly Caucasian environment exhibited an advantage in their recognition of Caucasian faces relative to their recognition of Korean faces. Experience with other races and a lack of further experience with own-race faces in early childhood can then reverse the other-race effect in recognition. Moreover, ethnicity is more than face and language!

References

- Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and nurture in own-race face processing. *Psychological Science*, 17 (2), 159-163.
- Bonatti, L., Frot, E., Zangl, R., & Mehler, J. (2002). The human first hypothesis: Identification of conspecifics and individuation of objects in the young infant. *Cognitive Psychology*, 44(4).
- Bowlby, J. (1969). Attachment and loss: Attachment (Vol. 1). New York: Basic.
- Bushnell, I. W. R., Sai, F. and Mullin, J. T. (1989). Neonatal recognition of the mother's face. *British Journal of Developmental Psychology*, 7, 3-15.
- De Haan, M., Johnson, M. H., Maurer, D., Perrett, D. I. (2001). Recognition of individual faces and average face prototypes by 1- and 3-month-old infants. *Cognitive Development*, 16, 659-678.
- Fantz, R. L. (1964). Pattern vision in newborn infants. *Science*, 140, 296-297.
- Field, T. M., Cohen, D., Garcia, R., Greenberg, R. (1984). Mother-stranger discrimination by the newborn. *Infant Behavior and Development*, 7, 19-25.
- Goren, C., Sarty, M. & Wu P.Y.K. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, 56, 544-549.
- Johnson, M.H., Dziurawiec, S., Ellis, H., Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40, 1-19.
- Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L. & Pascalis, O. (2007) The Other-Race Effect Develops During Infancy: Evidence of Perceptual Narrowing. *Psychological Science* 18: 1084-1089.
- Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., Ge, L., & Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, 8, F31-F36.
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., Iverson, P., Tsao, F. M. & Liu, H. M. (2006) Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science* 9: F13-F21.

- Lewkowicz, D. J. & Ghazanfar, A. A. (2006) The decline of cross-species intersensory perception in human infants. *Proc Natl Acad Sci U S A* 103: 6771-4.
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias memory for faces: A meta-analytic review. *Psychology, Public Policy & Law*, 7, 3-35.
- Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant and Child Development*, 10 (3), 3-18.
- Pascalis O., & de Schonen, S. (1994). Recognition memory in 3-4-day-old human infants. *NeuroReport*, 5, 1721-1724.
- Pascalis, O., de Haan, M., & Nelson, C.A. (2002). Is Face Processing Species-Specific During the First Year of Life? *Science*, 296, 1321-1323.
- Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & Fabre-Grenet, M. (1995). Mothers' face recognition by neonates - a replication and extension. *Infant Behavior and Development*, 18, 79-85.
- Patterson, M. L. & Werker, J. F. (1999) Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior & Development* 22: 237-247.
- Poulin-Dubois, D., Serbin, L. A., Kenyon, B. & Derbyshire, A. (1994) Infants' intermodal knowledge about gender. *Developmental Psychology* 30: 436-442.
- Quinn, P. C., Slater, A. (2003). *Face perception at birth and beyond*. In Pascalis, O., Slater, A. (Eds.). The development of face processing in infancy and early childhood : Current perspectives. Huntington, NY: Nova Science Publishers.
- Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception*, 31, 1109-1121.
- Sai, F. Z. (2005). The role of the mother's voice in developing mother's face preference: evidence for intermodal perception at birth. *Infant and Child Development*, 14 (1), 29-50.
- Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*, 16, 440-444.
- Spelke, E.S. (1979) Infants' intermodal perception of events. *Cognitive Psychology*, 6;8:553-560.
- Turati, C., Cassia, V.M., Simion, F., & Leo, I. (2006). Newborns' face recognition: Role of inner and outer facial features. *Child Development* 77 (2): 297-311.).
- Valentine, T. (1991). A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition. *The Quarterly Journal Of Experimental Psychology*, 43A (2), 161-204.
- Valenza, E., Simion, F., Macchi Cassia, V., Umiltà, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance*, 22 (4), 892-903.
- Walker-Andrews, A.S., & Lennon, E. (1991). Infants' discrimination of vocal expressions: contributions of auditory and visual information. *Infant Behav Dev*. 14:131-142
- Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N. & Werker, J. F. (2007) Visual language discrimination in infancy. *Science* 316: 1159
- Werker, J. F. & Tees, R. C. (2005) Speech Perception as a Window for Understanding Plasticity and Commitment in Language Systems of the Brain. *Developmental Psychobiology. Special Issue: Critical Periods Re-examined: Evidence from Human Sensory Development* 46: 233-234.

Is speech special?

Casey O'Callaghan
Rice University, Philosophy Department

There is a thriving debate over what aspects of our capacity to produce and understand language are special. My concern here is a key part of this wider debate: Is speech special? In particular, my focus is on speech perception, and whether it is special. This isn't just one but a number of different questions. Too frequently, these very different questions are not clearly distinguished and kept apart. I discuss a framework for distinguishing various versions of the question, Is speech perceptually special? Focusing on a particular class of questions, I make a proposal about the sense in which speech is perceptually special. According to this account, the capacity to perceive speech is an acquired perceptual skill, and involves learning to hear language-specific types of biologically-significant sounds. This account illuminates the significance of interlocution in understanding what makes the perception of speech distinctive.

1 Is speech perceptually special?

There is a thriving debate over whether the faculty of language is special (see, e.g., Hauser et al. 2002; Pinker and Jackendoff 2005). The question is if our capacity to produce and understand language is special. A key part of this wider debate is whether speech is special (see, e.g., Liberman 1996; Trout 2001). My concern here is speech *perception*. Is speech perception special?

Of course, answering this question means answering not just one but a number of different questions. Too frequently, these different questions are neither clearly distinguished nor kept apart. My goal is to present a framework for distinguishing various versions of the question, Is speech perception special? Focusing on a particular class of questions as an illustration, I make a proposal about the sense in which speech is perceptually special. This account illuminates the significance of interlocution for understanding what makes the perception of speech distinctive.

2 Many questions

To be special requires at least a *difference*. This raises the question, How different? The debate about whether speech is special aims at a stronger claim than that perceiving speech somehow differs from other perceptual capacities. Usually, it concerns whether speech perception is *distinctive*, or can be distinguished as a distinct variety among other forms of perception. Often, the question considered is stronger yet. Is speech perception *unique*, or the only instance of its kind?

Put this way, the question relies on a comparison. The most common contrast is with *general audition*. Is speech perception to some degree different from or unique in contrast to *non-linguistic* (or “ordinary”) *audition*? A separate contrast (too often used interchangeably with the first) is to the capacities of *non-human animals*. Are humans alone in having the capacity to perceive speech? Is speech perception uniquely human?¹

Furthermore, a difference is a difference in some respect, and being distinctive or unique is being distinctive or unique in some way or for some reason. So the most important question is, In what respect is speech perception special? Here there are a number of candidates. I find it helpful to divide them into two broad classes.

The first class of questions (Type 1, or “What”, questions) are forms of the question, *What* do we

1 Another contrast might be made between speech perception and *perception in general*. Is perceiving speech different in kind from or unique among varieties of perception? Finally, we might consider the contrast between *perceiving* speech and *understanding* speech.

perceive when we perceive speech? This first class includes questions about the *phenomenology*, *objects*, and *contents* of speech perception. Is there something special about the phenomenology, or the experience, of perceiving speech? Does speech perception have special objects? Does the perception of speech involve special contents?

The second class of questions (Type 2, or “How”, questions) are forms of the question, *How* do we perceive speech? Or, What are the *mechanisms* of speech perception? This second class includes questions about the *processes*, *module*, or *modality* involved in speech perception. Does perceiving speech involve special perceptual processes? Does speech perception involve a special perceptual module? Is speech perception itself a special perceptual modality?

The question, Is speech special?, thus becomes: Is human speech perception different, distinctive, or unique, when compared to non-linguistic audition, or to the perceptual capacities of non-human animals, in respect of its phenomenology, objects, or contents, or in respect of the processes, modules, or modality it involves?

I won’t tackle all of this here. Instead, I’ll focus on the Type 1 questions, which concern *what* we perceive when we perceive speech. Part of the reason for this is historical. There is a history of using answers to *what* questions to ground claims about *how* we perceive speech. Liberman (see, e.g., 1996) famously argued that the objects of speech perception differ from the objects of non-linguistic audition, and used this to argue that speech perception and non-linguistic hearing involve different perceptual modalities. This approach makes sense. We need to be clear about what the task is in order to understand what is required to perform it. So, it is particularly important to get the answers to Type 1 questions right.

3 Phenomenology

Is perceiving speech phenomenologically special? Is what it’s like for the subject to perceptually experience speech different or unique in relation to non-linguistic audition?

It’s common to think that the perceptual experience of speech is phenomenologically distinctive. Introspectively, there is a *perceptual* phenomenological difference between the experience of listening to speech and listening to non-speech sounds. But it can be difficult to motivate a phenomenological claim. Since we are good at detecting phenomenological contrasts, that is a good way to start.

Consider the contrast between listening to non-linguistic sounds and listening to spoken language. First, take the case of a language you know. It strikes me that there is a significant qualitative difference between the experience of listening to a language I know and the experience of listening to non-linguistic environmental sounds. Sinewave speech seems to confirm this. The same stimulus first is experienced as non-speech sounds, and then it is experienced as speech. However, in this case you *comprehend* the speech. Understanding might suffice to explain the phenomenological difference in listening to speech in a language you know. You grasp meanings, so the experiential difference could be explained in terms of cognitive, rather than perceptual, phenomenology.

So, consider listening to non-speech sounds in contrast to listening to speech in a language you do not know. Is there a perceptual phenomenological difference? It *seems* to me that there is some difference. The fact that neonates perceptually prefer speech sounds from non-speech sounds hints that there’s a difference, since neonates do not yet understand language. But reflection and babies are inconclusive here. Is the difference due to a difference in strictly audible characteristics of the stimulus, or does some further perceptual phenomenological difference accrue in virtue of its seeming like speech? If you could experience sinewave speech in a language you don’t know either as non-speech sounds or as speech sounds, that would provide good evidence for a perceptual phenomenological difference.

But consider the contrast between listening to speech in a known language and listening to speech in an unknown language. To control for differences in the stimulus, make it the same language. So, consider either the experiences of one person listening to speech in some language before and after learning the language, or consider the experiences of two similar listeners, one who knows the language and the other who doesn’t. While in each case there is a cognitive difference, that does not suffice to capture the phenomenological difference. There now also is good evidence for a perceptual phenomenological difference. While the stimulus is the same, perceptually experiencing speech in a language you know differs in a couple of respects. First, and most obviously, it has different temporal characteristics. You hear (and perhaps exaggerate) gaps and pauses, and better resolve temporal features

and contrasts. And, it has different qualitative characteristics. You detect subtle changes, contrasts, and differences in qualitative features you couldn't hear before. The stimulus *sounds* different when you recognize it as speech and know the language.

This all suggests that there is a *perceptual* (and not merely cognitive) phenomenological difference between listening to speech in a language you know and listening either to speech in a language you do not know or to non-speech. So there's good reason to think that speech perception has a distinctive phenomenology compared to non-linguistic audition.

4 Objects

The perceptual phenomenological difference might suggest that *what* you perceive when you perceive speech differs from what you hear when you listen to non-linguistic sounds. One kind of difference in what's perceived is a difference in the *objects* of perception. So, the phenomenological difference might suggest that the objects of speech perception differ from those of non-linguistic audition. In fact, researchers commonly have claimed that the objects of speech perception differ from those of audition generally. Evaluating the suggestion requires asking about the objects of speech perception and audition.

What are the intentional objects of speech perception? One primary focus of research into speech perception has been upon *phonemes*, whose patterns form the basis for recognizing and distinguishing words.

What is a phoneme? Phonemes are the minimal linguistically significant ways in which spoken words in a language differ perceptually. They're like the basic perceptible vocabulary that comprises spoken words in a language. For instance, the spoken word 'bad' includes /b/, /æ/, and /d/, while 'bat' and 'bash' differ because the former contains /t/ and the latter contains /ʃ/.

Phonemes are language-specific. *Phones*, on the other hand, include all of the perceptually discernible differences that can be linguistically significant in human languages. Phones are individuated in terms of the auditorily discernible (humanly producible) differences that could be exploited by a spoken language to signal a linguistically significant difference. Thus, phones are *types* that comprise auditorily equivalent perceptual objects. If audition's objects are sounds, phones are perceptually equivalent sound types. Phonemes, then, are classes of phones that, even though they might strictly speaking be perceptually distinguishable, are treated *as equivalent* within the context of a given spoken language. Hearers need not decide to treat strictly discernible sounds as members of a common type. When listening to a known spoken language as such, the listener cannot help but treat different phones as *allophones*, or as instances of a common phoneme. They are perceived as the same. On a natural picture, phonemes are perceptual equivalence classes of phones. Thus, the phonemes you perceive in perceiving speech are auditory equivalence classes of sounds. Phonemes are *sound types*.

Do the objects of speech perception differ from those of ordinary non-linguistic audition? Speech construed as such involves different *kinds of sounds* from non-linguistic environmental sounds, such as those of clucks, beeps, doors closing, cars backfiring, hands clapping, and dogs barking. But, according to the proposal above, the objects of speech perception and auditory perception belong to the same *ontological* kind, since both are types of *sounds*.

Nevertheless, the claim that speech perception and audition have different objects traditionally isn't just the claim that they involve hearing different *kinds of sounds*. It is the claim that perceiving speech is perceiving some object of an entirely different ontological kind from ordinary sounds. It is perceiving non-sounds. Many have argued that speech perception's objects do differ from non-linguistic audition's in this stronger sense. Three main sorts of argument are offered.

The strongest appeals to the *mismatch* between salient aspects of the experience of speech and features of the acoustic signal. The *acoustic* features that ground a perceived phoneme are highly context dependent. Not only do they vary in expected ways, with speaker, mood, and accent, but they also depend more locally upon the surrounding phonemes. Thanks to the effects of *coarticulation*, information about a given phoneme is blended with information about surrounding phonemes, and differs depending on the adjacent phonemes. While we experience /du/ and /di/ to share the sound of /d/, there is no invariant acoustic signature that corresponds to the /d/. Furthermore, while we experience a speech stream to be *segmented* into discrete phonemes and words, the acoustic information that cues /æ/ in 'dab' is present

during the articulation of both the /d/ and the /b/. There are no clear acoustic boundaries that correspond to those between experienced phonemes. In sum, there is no consistent, context-independent homomorphic mapping between experienced phonemes and straightforward features of the acoustic signal.

In light of this, Liberman and other proponents of the Motor Theory suggested that speech perception's objects are not sounds at all, but instead are aspects of the articulation of speech. The idea is that the gestures involved in the production of speech do map in a homomorphic, invariant way onto perceived speech. For instance, pronouncing /d/ involves stopping airflow by placing the tongue at the front of the palate, then releasing it while activating the vocal folds. Such gestures, and the features that combine to make them, make intelligible the perceptual individuation of speech in a way that the acoustic signal does not. The objects of speech perception and of audition thus differ in kind. The former are gestures, the latter are sounds.

This argument fails to establish that the objects of speech perception differ from the objects of non-linguistic auditory perception. On one hand, it relies on the premise that ordinary audition *does* map in an invariant and homomorphic way onto straightforward features of the acoustic stimulus. Even a simple audible quality like pitch has a complex relationship to frequency. And, especially in acoustically complex environments, the things we hear do not map straightforwardly and in an invariant way onto acoustic features. Context effects abound. For instance, the timbre of an enduring sound will differ if its attack differs. Furthermore, a central lesson of work on *auditory scene analysis* is that ordinary sounds are individuated—they are distinguished from each other at a time, and they are tracked and segmented over time—in the face of highly complex acoustic information (e.g., Bregman 1990). Nothing obvious in an acoustic stream signals how to distinguish the sound of a guitar from the sound of a voice in a crowded bar.

On the other hand, it relies on the premise that ordinary audition's objects *do not* map in an illuminating way onto aspects of the events and happenings that produce acoustic information. However, the sounds we hear are individuated in terms of features of their sources. This is reflected in how we talk about sounds: the sound *of the car door*, the sound *of the dog*, the sound *of scratching*. We carve up the auditory scene in large part in terms of sound sources and happenings. While the individuation of the objects of speech perception is illuminated by considering aspects of the articulatory gestures involved in speaking, this mirrors the fact that the individuation of the objects of non-linguistic auditory perception is illuminated by considering aspects of the happenings that make sounds. The function of auditory perception is to make perceptually accessible environmentally significant events. Sounds, among the objects of audition, are individuated not just in terms of the simplest physical properties of an acoustic stimulus, but in a way that facilitates awareness of sound sources. In this respect, speech perception does not differ in kind from non-linguistic audition. The *mismatch* argument fails.

Second, some argue that *cross-modal influences* in the perception of speech, such as the McGurk effect, reveal a substantial difference in the objects of speech perception and the objects of ordinary audition (see, e.g., Trout 2001 for discussion). Does the fact that visual information can impact which phoneme you experience show that, unlike sounds, which you can only hear, the objects of speech perception are available to both vision and audition? But in this respect speech is not unique. Cross-modal illusions and influences are rampant. Vision impacts non-linguistic audition (ventriloquism), ordinary audition impacts vision (sound-induced flash), vision alters touch (visual capture), touch alters audition, and so on (see, e.g., Spence and Driver 2004). Explaining each requires positing shared items among the perceptual objects of different modalities if the speech case does (O'Callaghan 2008). In fact, cross-modal effects support conceiving of audition's function as revealing the things and happenings that make sounds. Multimodality is not unique to speech.

Third, speech perception is *categorical*. That is, gradually varying a physical acoustical parameter leads to uneven perceptual variation. Where the effect is quite exaggerated, varying a physical parameter gradually might, for instance, cause one first to experience /b/ and then abruptly to experience a /d/ with little change noticed in between. In a dramatic case, an “analog” signal might cause apparently “digital” perception. Some have argued that the categoricity of phoneme perception means its objects are something other than ordinary sounds (see, e.g., Trout 2001; Pinker and Jackendoff 2005 for discussion). But, we now know that speech perception is not alone in being categorical. For instance, color perception is categorical, other forms of audition are categorical, and non-human animals show evidence of

perceiving categorically (see Harnad 1987; Cohen and Lefebvre 2005). The categoricity of phoneme perception cannot ground an argument that the perceptual objects of ordinary audition and speech perception differ in ontological kind.

If no good argument shows that the objects of speech perception and ordinary audition belong to entirely different kinds, what is the difference between perceiving speech and perceiving ordinary environmental sounds? If speech is just a variety of sounds that are perceptually individuated in terms of features of their sources, what explains the perceptual phenomenological difference between hearing speech and hearing ordinary sounds?

5 Contents

What we perceive when we perceive speech does, in another sense, differ from what we perceive when we perceive non-speech sounds. Speech perception has different *content* from non-linguistic audition. Content concerns how things are represented to be, or what features objects are perceptually attributed. One gloss on the contents of perception is in terms of the accuracy or veridicality conditions of a perceptual state. Another way to put this is in terms of what a given experience purports to be facts about the world (see Siegel 2005). My claim is that the content of speech perception differs from that of non-linguistic audition in two noteworthy respects.

First, the audible features of a stimulus perceptually experienced as speech differ from those of non-speech sounds. Experiencing speech involves experiencing certain fine-grained qualitative and temporal details that one does not normally experience when listening to even the same stimulus as non-speech. Speech experience discerns more and different *qualitative* details and contrasts than non-linguistic audition. Speech also audibly appears to have different and finer-grained *temporal* features—we hear gaps and pauses we didn't hear before. Further, we *segment* the sound stream differently over time when we hear it as speech. The individuation of sounds in time differs when we hear those sounds as speech. What formerly sounded like a continuous babble comes to sound like discrete sounds, syllables, and words. If sounds are audible individuals, we *hear different sounds* when we hear speech.

The second difference stems from the fact that human speech perception is *categorical*. While physical acoustical features vary along a continuum, phonemes are language-specific classes defined by perceptual equivalence of a certain sort. Belonging to a given phoneme category is an all-or-nothing matter, so perceiving phonemes is a kind of classificatory perception (cf. Matthen 2005). What consequences does this have for characterizing the content of speech perception? Perceiving speech is hearing sounds in a way that is consistent with their belonging to language-specific categories. Since these categories are equivalence classes of sounds, perceiving speech involves hearing sounds to stand in certain relations of similarity and difference to each other. These patterns of similarity and difference form a speech-specific (and language-specific) similarity space among sounds, in which regions correspond to particular language-specific speech sounds such as phonemes. In fact, one way to characterize these speech-specific categories is in terms of the speech-specific similarity relations among sounds. In hearing speech, the features sounds are perceived to have match this speech-specific similarity space among sounds. The content of speech perception, which grounds its difference from non-linguistic audition, reflects a distinctive pattern of similarity and difference among sounds.

6 How questions

What are the implications for questions about *how* we perceive speech? Type 2 questions concern the *means* or *mechanisms* involved in speech perception. Specifically, they ask whether speech perception involves special processes, a special module, or a special modality.

The evidence strongly suggests perceiving speech involves at least some special perceptual processes. Duplex perception for dichotic stimuli, developmental (especially critical period) differences, brain activity revealed by functional imaging, and dissociated disorders for speech and auditory perception all provide evidence of processes devoted to the perception of speech (see, e.g., Trout 2001).²

² Nevertheless, we should take care. Rich physiological and functional connections exist between general auditory and multimodal areas and language-specific areas. So, it is not always entirely clear whether some activity or

But, perhaps we can make do with a *minimal* story about the sense in which speech perception is special without appealing to special modules or even a special modality devoted to speech perception.³

This story is framed in terms of our *treatment* of speech and speech sounds. It involves two main claims, which are drawn from facts that must be accommodated by any adequate contemporary account of speech perception.

First, humans have a special or differential *selectivity* or *sensitivity* for the sounds of speech, in general. The striking evidence is that neonates distinguish and prefer speech to non-speech (Vouloumanos and Werker 2007). The sounds of speech in general are special for us, and they receive different treatment from other kinds of environmental sounds.

How do infants perceive speech sounds if speech sounds comprise *language-specific* classes of sounds? Humans are not born with the capacity to perceive phonemes *as such*. Very young infants in fact discern phonological differences from all languages—they distinguish among all of the possible speech sounds their language could include. Later, between 5-9 months, infants discern only the phonetic differences relevant to their own language. The usual story is that infants prune or forget how to perceive audible differences among phones that are not significant in their own language. Humans perceptually learn to ignore differences that are irrelevant to their language. Doing so is learning to treat one's language's allophones as such while losing the ability to distinguish sounds that other languages count as distinct phonemes. Such learning alters the language-specific similarity space among sounds, so we come to perceive sounds in a way that is consistent with their belonging to the relevant language-specific equivalence classes. We learn to discern the language-specific classes of sounds that comprise our language's phonemes. So, second, humans have a propensity for learning to perceive *language-specific* sound types.

Perceiving speech sounds from a known language, according to this understanding, requires experience and learning. It is an *acquired* perceptual skill. One learns to *hear* the sounds of one's language. Thus, learning a language is not just a matter of learning a sound-meaning mapping. It involves acquiring the auditory skill of *hearing* sounds in a way consistent with their belonging to language-specific perceptual equivalence classes. Learning a language is partly a matter of learning a perceptual skill.⁴

Since the capacity to perceive speech sounds in accordance with language-specific categories is an acquired perceptual skill, it differs (at least in degree) in this respect from the capacities to perceive individuals such as three-dimensional objects and events, persistence, and sensible qualities like color, pitch, and loudness. Arguably, these are capacities humans possess much earlier. On the other hand, speech perception may be more like our capacities to perceive things like clapping hands, dog barking, metal scraping metal, or fingernails scratching a chalkboard. These are best understood as acquired perceptual capacities.

7 Interlocution

What is the role of interlocution according to this account? Hearing speech is an acquired perceptual skill for which we have a special propensity from before birth. It involves learning to perceive language-specific types or equivalence classes of sounds, whose individuation is illuminated by

process is perceptual or extra-perceptual. So it is not entirely clear whether perceiving speech involves special *perceptual* processes.

3 While I won't argue for it here, I'm reluctant to say that speech perception involves a distinctive or unique perceptual modality independent from ordinary audition. Whether we individuate modalities in terms of their objects, phenomenology, function, or physiology, the evidence doesn't require a separate modality to deal perceptually with speech. While speech certainly may be handled differently from non-linguistic sounds by audition, colors and objects also are handled in different ways by vision. Similarly, I doubt speech perception is accomplished by a devoted perceptual module. If a process is modular only if it is informationally encapsulated, then speech perception isn't a module. Appelbaum (1998) argues convincingly against Fodor that domain general top-down influences impact the perception of speech sounds. Further, as I've argued here, audition and speech perception to a significant extent share function.

4 It is no objection to the claim that perceiving speech is acquired or learned that perceiving speech requires a special propensity, which must be innate. So does walking on two legs.

considering salient happenings in the environment: articulatory gestures and talking faces. Considered as such, perceiving speech is a matter of detecting and discerning *biologically significant* kinds of sounds and happenings, rather than just detecting abstract features of an acoustic signal.

How does perceiving speech differ from perceiving other biologically significant sorts of environmental sounds? Consider a family of capacities that reveal varieties of *animacy*. For instance, we might perceive a pattern of moving dots as *running*, or one dot to *chase* another dot around a display (Heider and Simmel 1944; see also Scholl and Tremoulet 2000). Here we describe the perception of inanimate things and motion in terms applicable to animate things and activities on the basis of very minimal cues, which suggests we have a special propensity to perceive animate things and activities. Perceiving speech is similar to perceiving these other special sorts of biologically significant things and activities, in that its concern is a type of *animacy* exhibited by living things to which we have special sensitivity. That is, we have differential sensitivity to certain kinds of *activity* that creatures engage in, in contrast to simple motion patterns or inanimate happenings. Furthermore, in the case of speech, this capacity is directed at members of our own species, as is the capacity to perceive *faces*.

Speech sounds also belong to an even more special subclass because they are generated by *communicative intentions* of other humans. Like facial expressions and some non-linguistic vocalic sounds, speech sounds are caused by and thus have the potential to reveal the communicative intentions of their animate sources. Speech perception thus belongs to a special class of perceptual phenomena that serve to reveal biologically significant intentional activities involved in communication. Perceiving speech is detecting and discerning language-specific kinds of biologically significant events: ones that are generated by communicative intentions of fellow human talkers. We hear people talking. We hear them as interlocutors.

References

- Appelbaum, I. (1998). Fodor, modularity, and speech perception. *Philosophical Psychology*, 11(3):317–330.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge, MA.
- Cohen, H. and Lefebvre, C. (2005). *Handbook of Categorization in Cognitive Science*. Elsevier, New York.
- Harnad, S. (1987). *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, Cambridge, UK.
- Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298:1569–1579.
- Heider, F. and Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2):243–259.
- Liberman, A. M. (1996). *Speech: A Special Code*. MIT Press, Cambridge, MA.
- Matthen, M. (2005). *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception*. Oxford University Press, Oxford.
- O’Callaghan, C. (2008). Seeing what you hear: crossmodal illusions and perception. *Philosophical Issues*, 18:316–338.
- Pinker, S. and Jackendoff, R. (2005). The faculty of language: what’s special about it? *Cognition*, 95:201–236.
- Scholl, B. and Tremoulet, P. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8):299–309.
- Siegel, S. (2005). The contents of perception. In Zalta, E. N., editor, *Stanford Encyclopedia of Philosophy*.
- Spence, C. and Driver, J., editors (2004). *Crossmodal Space and Crossmodal Attention*. Oxford University Press, Oxford.
- Trout, J. D. (2001). The biological basis of speech: what to infer from talking to the animals. *Psychological Review*, 108(3):523–549.
- Vouloumanos, A. and Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Developmental Science*, 10(2):159–164.

Casey O'Callaghan
<http://ocallaghan.rice.edu>

Towards embedded and embodied accounts of language use: insights from an ecological perspective

Navin Viswanathan
University of Connecticut & Haskins Laboratories

One of the fundamental tenets of ecological accounts of cognition is that the organism-environment system cannot be meaningfully partitioned during investigation. Subsumed under this overarching assertion is the insight that organism-environment mutuality should also apply to components of a perception-action-cognition system. In this paper, I start by examining how insights from an ecological account have been applied to the study of language (specifically to the study of speech perception, e.g., Fowler, 1986). I then attempt to apply the insights in understanding two aspects of language use: embodiment and embeddedness. An embodied approach to language (and cognition more broadly) suggests that perception, action, and cognition are strongly interrelated and are served by the same systems. I examine recent findings of embodiment that appear to implicate the motor system in cognition and perception (see Galantucci, Fowler, & Turvey, 2006, for a review). An embedded approach to cognition, on the other hand, recognizes that any aspect of cognition cannot be fully described separately from the context in which cognition occurs. For language, I suggest that this embeddedness exists at two levels. First, language use typically occurs in a rich physical environment and incorporates the constraints of the physical world (e.g., Chambers, Tanenhaus, & Magnuson, 2004). Second, language users are embedded in a shared socio-cultural environment, which shapes and constrains their language use (e.g., Giles, Coupland & Coupland, 1991). I suggest that the insights drawn from ecological accounts of cognition, coupled with existing findings of embedded-embodied language use could provide a fresh theoretical framework that can be employed toward understanding language use.

1. Introduction

The ecological approach to visual perception (Gibson, 1966, 1979) takes a radically different approach to the study of perception as compared to more traditional computational accounts (e.g., (Neisser, 1967). In this paper, I lay out the fundamental assumptions of an ecological account, and examine how the application of even a subset of these principles could provide us with alternate frameworks for studying language.

2. Ecological principles

2.1 Assertion of realism

The first assertion is that there exists a real world and the primary task of the organism is to get to know and act successfully in this world. While this may seem like a truism for most accounts of perception (see, Shaw, Turvey, & Mace, 1982, for a discussion), an explicit recognition of this assertion influences what we regard as primary in an ecological study of cognition. Thus, rather than starting from typically studied “higher cognitive processes” such as thought, imagery or language to provide explanation for cognition, the ecological argument is that these capabilities must have been built upon existing perceiving-acting abilities (see Brooks, 1997). For the study of language, it would mean considering seriously the environmental and social interactions that language fosters.

2.2 Direct perception

The second principle of an ecological account is that perception is unmediated by internal representations or inferential processes. This claim constitutes the major point of departure from computational accounts of cognition, where the focus is mainly on what aspects of the world are recreated in the organism and how these internal recreations are manipulated to produce more abstract mental structures. From the ecological account, the focus is on how the organisms detect information provided by the environment to support their activities. Thus, their role is of information seekers and detectors, rather than information enrichers or creators as in traditional computational accounts of cognition (see Michaels & Carello, 1981). The assertion of direct perception hinges on the next assumption of informational specificity.

2.3 Informational specificity

This principle holds that there exists a one-to-one relationship between the environmental event to be perceived and the structure of the resulting informational array. Information can be carried through optical structure in reflected light (detected visually), acoustic structure in the form of compression waves (detected auditorily), chemical structure (detected through olfaction or gestation), etc. Specificity implies that a given event produces a unique structure in the informational array and given this structure, the organism can unambiguously detect the causative event by recruiting the appropriate sensory modality. Thus, the richness in behavior is attributed to the richness of stimulation, rather than elaborate rational processes that act on an impoverished signal (Michaels & Carello, 1981).

Adopting this subset of ecological principles can alter our approach to the study of language significantly. In the next section, I outline the direct realist account of speech perception as an illustration of how these principles have been applied.

3. The direct realist theory of speech perception (Fowler, 1986)

Fowler's account of speech perception embodies the principles outlined above. Broadly, according to this account, the process of speech perception proceeds as follows. During conversation, the speaker through carefully coordinated actions of the vocal tract creates the acoustic signal. The resulting acoustic signal contains information about how it was produced (viz. the sequence of vocal tract gestures that produced it). The listener's task is now to detect this information, in order to perceive the source of the signal – the vocal tract gestures. This is in stark contrast with computational accounts of speech perception, wherein the task of the listener is to search the acoustic signal for cues to activate relevant mental representations through appropriate statistical processes of inference (e.g. Diehl, Lotto, & Holt, 2004). In the direct realist account, speech perception proceeds without reference to mental representation or recruitment of inferential processes. Rather, the focus is on what properties of the acoustic signal are specific to the vocal tract gestures and consequently support speech perception.

The objects of speech perception by this account, like in a general ecological account, are events in the world. In the case of speech they happen to be gestures of the vocal tract. Listeners are able to detect vocal tract gestures from the acoustic signal (rather than infer what would have caused it) because of specificity. The outstanding challenge then for this account is to pinpoint the invariance in the acoustic signal.

On the issue of whether speech is special (see O'Callaghan, this issue) the direct realist account disagrees with the other prominent gestural account (The Motor theory, Liberman & Mattingly, 1985). By the direct realist account, listeners perceive speech and non-speech alike in that they have the same objects of perception: environmental events. Thus speech by this account (on the basis of its objects of speech perception) is not special but is like the auditory perception of any other event in the world.

The previous section outlined three ecological principles illustrating their application to the study of language by considering the direct realist speech perception account. Some of the principles may be seen as quite radical as compared to more traditional approaches to cognition and specifically language. In this paper, I suggest that even if we remain agnostic of these claims, our approaches may still gain substantially from considering seriously the following two principles in themselves in the study of language.

4. The interrelationship between perception, action and cognition

The interrelationship between perception and action is a key principle of the Gibsonian account. A perception-action cycle is composed of related perceptual and motor activity that constitutes a meaningful interaction with the environment. Cognition is viewed as grounded in perception-action cycles and is inherently coupled with perception and action (e.g., Brooks, 1997).

For instance, consider the concept of an *affordance* – what the environment offers the organism for good or for ill (Gibson, 1979). In other words, affordances are the action possibilities offered by the environment. The perception of affordance, which is of primary relevance in this account, is a case where this account stresses understanding perception in the context of supporting action. The challenge for the organism is to use *perceptually guided* action to interact appropriately with the environment.

Similarly, Gibson writes of perceptual systems requiring a series of deliberate, explorative actions for successful perception. Unlike most other approaches, the process of, for example, visual perception, is not explained by having a stationary eye and starting from the retinal image. Rather the visual perceptual system, by this account, consists of two eyes mounted on a head that is in turn mounted on a neck and involves numerous actions that involve scanning global properties of the optical array. So the *simple* act of perception involves a host of explorative actions including head turns, postural changes, etc. Such a case constitutes an example of action serving perception. In short, perception and action form an inseparable pair for the organism and studying them separately without considering the other is unlikely to provide an ecologically valid account of behaviour.

The argument I wish to make here is that the same is demonstrably true for studying linguistic behavior. Several findings from the embodied cognition perspective appear to demonstrate the inseparability of perception, action and cognition. Specifically, they demonstrate the involvement of the motor system in perception.

For instance, Glenberg and Kaschak (2002) demonstrated that participants were faster to respond if their response direction was compatible with the direction implied by the sentence than if the opposite were true. For instance (1) was responded to faster than (2) if the 'yes' response for a sentence sensibility judgment task consisted of moving the lever towards the participant.

- (1) Andy delivered you the pizza.
- (2) You delivered the pizza to Andy.

Similarly, Casassanto and Lozano (2007) showed that downward responses were facilitated while judging valence of negatively valenced words like 'gloomy'. Furthermore, Bargh, Chen, & Burrows (1996) showed that participants, after reading words like "Florida" or "Bingo" that are typically associated with the elderly, walked more slowly to the elevator than control subjects. While, the interpretation of these findings are typically in disagreement with the ecological principles outlined previously, they support the current claim under discussion (of perception-action-cognition interrelatedness).

4.1 Interrelationship between speech perception and production

This principle has important implications for the study of speech perception and production. Several findings appear to indicate strong perception-production links in speech. Here I will review a few such findings to demonstrate the kinds of questions that can stem from adopting such an approach.

a. The effect of perception on production. Kerzel and Bekkering (2000) showed that listeners were facilitated in the production of syllables when a distracting visual display contained gestures that matched those of the target syllables. Galantucci et al., (in press) extend this finding by demonstrating that such stimulus-target compatibility occurred when the distracting syllables were presented auditorily as well. In addition, Gentilucci and colleagues (Gentilucci, 2003; Gentilucci, Santunione, Roy, & Stefanini, 2004) demonstrate the effects of perceiving non-linguistic objects on speech production, hinting at a more general perception-production link.

b. The effect of production on perception. Several findings appear to reliably implicate the role of the motor system in speech perception. In typical McGurk studies (McGurk & MacDonald, 1976), a visually presented syllable, eg., /ba/, when combined with an auditorily presented syllable /ga/ is often reported to be heard as /da/ demonstrating influence of the visually presented speech. Sams et al., (2005) showed that the same McGurk-like influences resulted when instead of a visually presented discordant syllable, the listeners themselves silently uttered them, demonstrating that listener's own actions can alter perception. Similarly, Ito, Tiede, & Ostry (2009) showed that when listeners' skin was stretched (by a robotic device) in a manner similar to the stretching that occurs during speech production, listener's categorization of vowel sounds changed. Finally, in a study examining the same question (of the perception-production link), Viswanathan et al. (in prep) had subjects perform movements of the lip or the tongue while classifying a [ba] to a [da] continuum. One group of listeners performed these movements simultaneously while hearing the spoken syllable, while the other group performed these movements 500 ms before (lead condition) they heard the syllable.

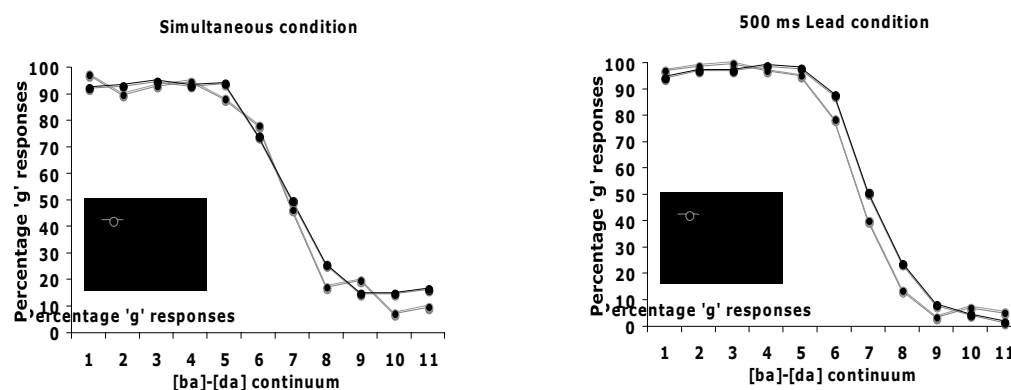


Fig 1. The effect of concurrent and preceding speech-like movements on stop categorization.

Figure 1, shows the results of this experiment. Listeners in the lead condition reliably reported fewer “b” responses while making lip movements than tongue movements demonstrating a form of interference in their phonetic judgment. No effect of listener movement was detected in the simultaneous condition.

These findings appear to support the idea that speech perception and speech production are closely linked. Apart from these behavioral studies, there also exists a rich literature using other techniques that demonstrate this link (see D’Ausillo, this issue for a detailed review and supporting findings). In general, it is clear that a thorough understanding of spoken communication would involve studying speech perception and speech production in the context of each other and grounded in an appropriate theory of phonology.

5. The organism-environment mutuality

This is arguably the most important of ecological principles and is directly relevant for accounts focused on describing language use in its proper context. Put simply, this principle is that the organism-environment system cannot be meaningfully partitioned during investigation. In other words, proper understanding of behavior results from studying cognition in its proper context. The phenomenon of perception, action and cognition, from this account is spread over the organism-environment system and consequently it is impossible to situate any cognitive process including language solely in the organism (in the human language user). By mistakenly partitioning the system during investigation, we may both complicate the particular phenomena under study as well as develop accounts of such phenomena that do not generalize to their natural contexts in which such behavior occurs.

By this principle, an account of language has to strive to describe linguistic behavior that is embedded in its natural contexts. For language I suggest that this idea of embeddedness occurs at two levels.

5.1 Physical embeddedness

First, the interlocutor is immersed in a physical environment that is rich in its structure. Part of understanding language use involves adequately describing how interlocutors utilize this rich structure to communicate (also see “common ground”, Clark, 1996). The issue of physical embeddedness places constraints on typical accounts of language perception and production to consider and incorporate information from the physical world in which language use is embedded. For instance, in the area of spoken word recognition, several studies use the visual world paradigm, which relies on listener’s ability to rapidly attend to aspects of the visual environment based on spoken utterances (e.g. Allopena, Magnuson, & Tanenhaus, 1998). This technique demonstrates that embedding language users even in simplistically structured worlds provides us with better tools for understanding language use. However, the issue of physical embeddedness can be pushed even further. Chambers et al., (2004) designed an ingenious set of experiments to ask whether the constraints of their physical environment affected how they parsed syntactically ambiguous sentences. Specifically they investigated whether the affordances of the objects in their physical environment altered their parsing of sentences. For instance, listeners were presented with sentences such as (3)

(3) Pour the egg in the bowl over the flour

The critical question being whether listeners temporarily consider the bowl as a possible destination for the egg. First, they demonstrated that when there was only one egg in the environment, listeners showed more looks to the bowl than when there were two eggs (where “in the bowl” disambiguated the referent). This finding already shows listeners rapidly incorporate properties of the physical world in their sentence comprehension. Next the experimenters compared looks to the distractor (bowl) when both eggs were placed in the bowl, but only one was pourable (the other was a solid egg). Interestingly, listeners behaved as though there was only one egg in their environment attuning to the different affordances of the two eggs. Finally, they demonstrated that when the affordances of the referents were modified by placing a tool in the listener’s hand, listeners in their sentence parsing incorporated the altered action possibilities in the presence of the tool. These kinds of findings demonstrate the importance of the physical environment in language use.

5.2 Socio-cultural embeddedness

In addition to the physical environment, language users are also embedded in their socio-cultural environment. For instance, language perception and production is constrained not just by aspects of the physical environment but also by the abilities of the interlocutor (e.g. Hanna & Tanenhaus, 2004). Furthermore, language users demonstrate convergence to aspects of their interlocutors linguistic and non-linguistic behavior (Giles et al., 1991; Fussell & Kraus, 1992; Chartrand and Bargh, 1999). Moreover, it has been shown that speakers converge to the phonetic characteristics of their environment (e.g., Sancier and Fowler, 1997) suggesting that language use is sensitive to the ambient social environment. Socio-cultural factors such as socio-economic status, attitudes, regional affiliation, sexual orientation etc., systematically affect several aspects of linguistic behavior (e.g., Labov, 1966) suggesting the importance of these factors in understanding linguistic behavior. Consideration of socio-cultural embedding would, for instance, alter our approach to speech perception.

Accounts of speech perception and production tacitly assume *parity* (Lieberman & Whalen, 2000) – the notion that listeners and speakers have a common understanding of what sounds are linguistically relevant (and what these sounds are). If we are to take the fact that interlocutors are embedded in a rich socio-cultural environment seriously, then spoken language accounts would have to describe how parity is established and perhaps constantly negotiated among interlocutors in their continuously varying cultural environments. In short, I suggest that the insights from socio-linguistics must be considered seriously if one were to provide an ecologically valid account of language use.

6. Conclusion

In conclusion, the ecological principles, especially of perception-action-cognition interrelatedness and organism-environment mutuality, constitute suitable starting points for expanding our current understanding of language use. Whether or not one subscribes to all the principles of this approach, adopting a subset of these principles is still likely to considerably impact current approaches and accounts of language use.

References

- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419-439.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71, 230-244.
- Brooks, R. A. (1997). *Intelligence without representation*. Cambridge, MA: MIT Press.
- Casasanto, D., & Lozano, S. (2007). Meaning and motor action. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 149-154).
- Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3), 687-696.
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893-910.
- Clark, H. H. (1996). *Using Language*. Cambridge University Press.
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech Perception. *Annu. Rev. Psychol.*, 55, 149-79.
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14(1), 3-28.
- Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, 62(3), 378.
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361-377.
- Gentilucci, M. (2003). Grasp observation influences speech production. *European Journal of Neuroscience*, 17, 179-184.
- Gentilucci, M., Santunione, P., Roy, A. C., & Stefanini, S. (2004). Execution and observation of bringing a fruit to the mouth affect syllable pronunciation. *European Journal of Neuroscience*, 19(1), 190.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin Boston.
- Gibson, J. J. (1979). *The Ecological approach to visual perception*. Lawrence Erlbaum.
- Giles, H., Coupland, N., & Coupland, J. (1991). 1. Accommodation theory: Communication, context, and consequence. *Contexts of accommodation: Developments in applied sociolinguistics*, 1.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558-565.
- Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: evidence from eye movements. *Cognitive Science*, 28(1), 105-115.
- Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences*, 106(4), 1245.
- Kerzel, D., & Bekkering, H. (2000). Motor activation from visible speech: evidence from stimulus response compatibility. *Journal of Experimental Psychology: Human Perception and Performance*, 26(2), 634-647.
- Labov, W. (1966). The Social Stratification of English in New York City.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1-36.
- Lieberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4(5), 187-196.

- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746-748.
- Michaels, C. F., & Carello, C. (1981). *Direct perception*. Prentice-Hall Englewood Cliffs, NJ.
- Neisser, U. (1967). *Cognitive Psychology*. Prentice-Hall, New Jersey.
- Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Brain Research. Cognitive Brain Research*, 23(2-3), 429-435.
- Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25(4), 421-436.
- Shaw, R. E., Turvey, M. T., & Mace, W. (1982). Ecological psychology: The consequence of a commitment to realism. *Cognition and the symbolic processes*, 2, 159-226.

Motor contribution to speech perception

Alessandro D'Ausilio

DSBTA – Section of Human Physiology, University of Ferrara, Italy

Classical models of language localization in the brain consider an antero-posterior distinction between perceptual and productive functions. In the last 15 years, this dichotomy has been weakened because of empirical evidence suggesting a more integrated view of action and perception mechanisms. Here we report some new data showing the role played by the motor system in speech perception. Specifically, we demonstrated that the motor system might be causally involved in the discrimination of degraded or noisy speech stimuli.

One of the experimental and theoretical streams that strongly opposes a strict perception-production dichotomy is the “motor theory of speech perception” (Liberman et al., 1967; Liberman and Mattingly, 1985; Liberman and Whalen, 2000). Among several other propositions, it claimed that the ultimate constituents of speech are not sounds but intended articulatory gestures. Speech perception and speech production processes use a common repertoire of motor primitives. In other words, the listener understands the speaker when his/her articulatory gesture representations are activated through verbal sounds. Although several aspects of this proposal have been modified in the past 40 years (Galantucci et al., 2006), interest in these aspects of the motor theory of speech perception has been recently revived by a series of neurophysiological experiments in the motor systems of monkeys. In a monkey's premotor area F5, hand and mouth actions are represented with a high degree of specificity. Neurons in this area discharge during grasping, holding, tearing or manipulating, whereas they are silent when the monkey performs actions that involve a similar muscular pattern but are driven by a different goal (Matelli et al. 1985, Rizzolatti et al. 1988). Among neurons of area F5, a subset of visuomotor neurons show even more interesting properties. Mirror neurons are active both during observation and execution of the same grasping action (di Pellegrino et al., 1992, Gallese et al., 1996; Rizzolatti et al., 1996a). Canonical neurons are active both during object hand-grasping and during observation of graspable objects (Rizzolatti and Fadiga, 1998; Murata et al., 1997; Jeannerod et al., 1995). Thus, the monkey premotor cortex may transform visual information into motor knowledge (Rizzolatti & Craighero, 2004).

These functional properties of mirror and canonical neurons found in the macaque premotor cortex strongly resemble the mechanism, though not the primary effectors, proposed by Liberman for speech perception. Every time the monkey observes the execution of an action or an object upon which actions might be organized, the related F5 neurons are activated and the specific action representation is “automatically” evoked. Under certain circumstances action representation may remain an unexecuted representation to be used in understanding what others are doing or categorize manipulable objects around us. The observation of actions done by others and/or of graspable objects activates, in humans as well, a cortical network including regions characterized by motor functions. The rostral part of the inferior parietal lobule (IPL), the ventral premotor area (vPM) and the pars opercularis of the inferior frontal gyrus (IFG) (Rizzolatti et al. 1996b; Grafton et al. 1996; Decety et al. 1997; Hari et al., 1998; Iacoboni et al. 1999; Binkofski et al., 1999; Chao and Martin 2000; Buccino et al., 2001; Nishitani and Hari, 2000; Grèzes et al. 2003). These regions form the core of the human mirror system. What is more interesting here is that pars opercularis of the IFG belongs to Broca's region (see Amunts et al., 1999), an area classically considered as the frontal center for speech production. It has to be stressed that, from a cytoarchitectonical point of view, human Broca's area closely resembles monkey premotor area F5 (Petrides, et al., 2005).

The first empirical demonstration of the validity of Liberman's theory comes from a TMS study demonstrating that when listening to speech, listeners' motor centers representing tongue movements motorically “resonate” as if they were actually producing the perceived speech (Fadiga et al. 2002). In their study, the authors recorded tongue motor evoked potentials (MEPs) elicited by TMS of tongue

motor cortex, while participants were listening to words, pseudowords and control sounds. Results showed that while listening to words and pseudowords formed by consonants implying tongue mobilization (i.e. Italian 'R' vs. 'F'), tongue motor potentials evoked by TMS were significantly increased. This indicates that when an individual perceives verbal stimuli, speech related motor centers are specifically activated. Since then, a large amount of data has been accumulated challenging the theory of a strict anatomo-functional segregation between speech perception and production mechanisms. Passive perception of phonemes and syllables activate motor (Fadiga et al., 2002; Watkins et al., 2003; Pulvermüller et al., 2003; Pulvermüller et al., 2006) and premotor areas (Wilson et al., 2004). Interestingly, these activations were somatotopically organized according to the effector recruited in the production of these phonemes (Fadiga et al., 2002; Watkins et al., 2003; Pulvermüller et al., 2006), and in accordance with the premotor activities in overt production (Pulvermüller et al., 2006; Wilson et al., 2004).

However, a distinctive feature of action-perception-theories in general, and in the domain of language specifically, is that motor areas are considered necessary for perception. All the above mentioned studies are inherently correlational, and it has been argued that in absence of a stringent determination of a causal role played by motor areas in speech perception, no final conclusion can be drawn in support of motor theories of speech perception (Toni et al., 2008). In fact, the mere activation of motor areas during listening to speech might be caused by a corollary cortico-cortical connection that has nothing to do with the process of comprehension itself. Therefore, a possible solution might come from the selective alteration of neural activity in speech motor centers and the evaluation of effects on perception. Meister et al., (2007) recently did a repetitive TMS study suggesting that vPM may play a role in phonological discrimination. In our view, however, this study fails to offer a convincing proof of the causal influence that motor areas may exert. Because of the spread and the variety of possible effects elicited by a 15 min TMS stimulation, such an offline rTMS protocol might have indeed modified the activity of a larger network of areas, possibly including posterior receptive language centers (Matsumoto et al., 2004). Moreover, there is no evidence of an effector-specific effect, i.e., that stimulating tongue representation induced specific deficits in the perception of tongue-related phonemes.

Therefore we designed a series of TMS experiments to tackle the causal contribution of motor areas to speech perception (D'Ausilio et al., 2009). In the first experiment we studied the role of the motor cortex in the discrimination of phonemes produced with the tongue or the lips (lip-related: [b] and [p]; tongue-related: [d] and [t]). On-line TMS pulses were applied either to the lip or tongue motor representations in motor cortex just prior to stimuli presentation. We found that focal stimulation facilitates the perception of the concordant phonemes, and inhibits perception of the discordant items. TMS stimulation on the motor representation of a given articulator (i.e. Tongue) prepares the system to perceive sounds produced with that specific effector ([d] and [t]), and interferes with other classes of speech sounds ([b] and [p]). Auditory stimuli were immersed in white noise so that recognition performance was kept around 75% accuracy. In the second Experiment, we replicated all conditions and parameters of the first experiment except that we removed the white noise. Stimuli were recognized with approximately 100% accuracy. Not surprisingly the effect we observed in experiment 1 disappeared.

Several studies verified the recruitment of the motor system when dealing with degraded speech signals. For instance Binder et al. (2004) had subjects identify speech sounds masked by varying levels of noise. Accuracy and response times were used to characterize the behavior of sensory and decision components of this perceptual system. Signals in the IFG predicted response time whereas temporal regions accuracy. They show evidence for a functional distinction between sensory and decision mechanisms underlying auditory object identification. Moreover Boatman and Miglioretti (2005) designed an intra-cortical stimulation study in which patients were divided in two groups according to their speech discrimination abilities. All of them were asked to perform a speech discrimination task while interfering brain stimulations were applied to several anterior and posterior language areas. The low-performance group also showed performance impairment when stimulated in frontal language areas, beside temporal cortices. Finally a recent study by Sato et al. (2009) showed that rTMS applied over the vPM resulted in slower phoneme discrimination when phonemic segmentation was required. No effect was observed in phoneme identification and syllable discrimination tasks that could be performed without need for phonemic segmentation.

In conclusion, our data together with other related studies suggest that the motor cortex

contribution might be critical only when faced with degraded acoustic stimuli or when the task require a relatively high degree of cognitive load. More specifically, the auditory system might have enough information for the successful discrimination of speech sounds when listening conditions are near perfect. However, these listening conditions rarely occur and rather we are typically exposed to ambiguous material, due to inter-speaker variability or external interferences. The motor system may be critical in these specific and ecological situations. This theoretical proposal is based on the idea that during development a sensory motor map is built via spontaneous vocal production followed by somatosensory and auditory feedback. From then on, feeding auditory information in the system may thus pre-activate motor commands to produce those sounds. When we listen to degraded speech it is still possible to extract some auditory information. Those features, although not sufficient for reaching a final classification, might pre-activate a subset of compatible motor candidates. These motor candidates in turn come with all the associated auditory features that are missing in the signal. Therefore, the activation of a subset of sensory-motor couples may be used to fill gaps in the incoming information or prime the detection of other auditory features. Through an iterative process, auditory information might select out motor candidates that may further direct auditory search in a lower space of features, thus facilitating and speeding up classification. Motor centers might therefore disambiguate or fill the gaps in the auditory stream via anticipatory mechanism.

References

- Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H. B., & Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *J Comp Neurol*, 412(2), 319-341.
- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci*, 7(3), 295-301.
- Binkofski, F., Buccino, G., Posse, S., Seitz, R. J., Rizzolatti, G., & Freund, H. (1999). A fronto-parietal circuit for object manipulation in man: evidence from an fMRI-study. *Eur J Neurosci*, 11(9), 3276-3286.
- Boatman, D. F., Miglioretti, D. L. (2005). Cortical sites critical for speech discrimination in normal and impaired listeners. *J Neurosci*, 25(23), 5475-5480.
- Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur J Neurosci*, 13(2), 400-404.
- Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage*, 12(4), 478-484.
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19, 381-385.
- Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., et al. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain*, 120 (Pt 10), 1763-1777.
- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Exp Brain Res*, 91(1), 176-180.
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur J Neurosci*, 15(2), 399-402.
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin Review*, 13, 361-377.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119 (Pt 2), 593-609.
- Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. *Exp Brain Res*, 112(1), 103-111.
- Grèzes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. *Neuroimage*, 18(4), 928-937.
- Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of

- human primary motor cortex during action observation: a neuromagnetic study. *Proc Natl Acad Sci U S A*, 95(25), 15061-15065.
- Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, 286(5449), 2526-2528.
- Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends Neurosci*, 18(7), 314-320.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol Rev*, 74(6), 431-461.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1-36.
- Lieberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends Cogn Sci*, 4(5), 187-196.
- Londei, A., D'Ausilio, A., Basso, D., Sestieri, C., Del Gratta, C., Romani, G., et al. (2007). Brain network for passive word listening as evaluated with ICA and Granger causality. *Brain Research Bulletin*, 72, 284-292.
- Londei, A., D'Ausilio, A., Basso, D., Sestieri, C., Del Gratta, C., Romani, G., et al. Sensory-motor brain network connectivity for speech comprehension. *Submitted to Hum Brain Mapp*.
- Matelli, M., Luppino, G., & Rizzolatti, G. (1985). Patterns of cytochrome oxidase activity in the frontal agranular cortex of the macaque monkey. *Behav Brain Res*, 18(2), 125-136.
- Matsumoto, R., Nair, D., LaPresto, E., Najm, I., Bingaman, W., Shibasaki, H., et al. (2004). Functional connectivity in the human language system: a cortico-cortical evoked potential study. *Brain*, 127, 2316-2330.
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, 17, 1692-1696.
- Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., & Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *J Neurophysiol*, 78(4), 2226-2230.
- Nishitani, N., & Hari, R. (2000). Temporal dynamics of cortical representation for action. *Proc Natl Acad Sci U S A*, 97(2), 913-918.
- Petrides, M., Cadoret, G., & Mackey, S. (2005). Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature*, 435(7046), 1235-1238.
- Pulvermuller, F., Shtyrov, Y., & Ilmoniemi, R. (2003). Spatiotemporal dynamics of neural language processing: an MEG study using minimum-norm current estimates. *Neuroimage*, 20, 1020-1025.
- Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences U S A*, 103, 7865-7870.
- Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Exp Brain Res*, 71(3), 491-507.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996a). Premotor cortex and the recognition of motor actions. *Brain Res Cogn Brain Res*, 3(2), 131-141.
- Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., et al. (1996b). Localization of grasp representations in humans by PET: 1. Observation versus execution. *Exp Brain Res*, 111(2), 246-252.
- Rizzolatti, G., & Fadiga, L. (1998). Grasping objects and grasping action meanings: the dual role of monkey rostroventral premotor cortex (area F5). *Novartis Found Symp*, 218, 81-95; discussion 95-103.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu Rev Neurosci*, 27, 169-192.
- Sato, M., Tremblay, P., Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain Lang*, doi:10.1016/j.bandl.2009.03.002
- Toni, I., de Lange, F. P., Noordzij, M. L., & Hagoort P. (2008). Language beyond action. *J Physiol Paris*, 102(1-3):71-79.
- Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41, 989-994.
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor

areas involved in speech production. *Nat Neurosci*, 7(7), 701-702.

When socio-linguistic interaction breaks down

Sonya Bird
University of Victoria

Usage-based models of language provide useful insights into how linguistic information is stored in the mind and accessed during speech production and perception. The crucial assumption made in these models is that members of a linguistic community are continually immersed in language, and this affects their linguistic behavior in specific ways. The question addressed in this paper is: how do usage-based models fare when accounting for patterns found in situations of restricted language use? To answer this question, two linguistic situations and two kinds of speech production data are considered: variability in the pronunciation of laryngealized resonants (phonetics) in St'át'imcets and variable use of consonant gradation (morpho-phonology) in Sointula Finnish. Based on these situations, it is proposed that usage-based models can account for patterns found in situations of restricted language use, as long as they include the flexibility to allow for 1) an expansion of what we mean by 'frequency effects' and 2) a way of including other effects, which can take over when frequency-related effects can no longer predict linguistic behaviour.

1. Introduction: Usage based models and language shift

Usage-based models of language (Bybee, 2001; Pierrehumbert, 2001; and others) have gained popularity in recent years because of their ability to incorporate a wide range of factors in a unified explanation of attested linguistic patterns. In the realm of sound, for example, they allow not only for purely linguistic (phonological) factors to affect speech outcome, but also socio-linguistic factors and factors relating to the physiological and cognitive attributes of speakers. The crucial basis of these models – as is clear from their name – is *usage*: the assumption is that members of a linguistic community are continually immersed in language, and this affects their linguistic behavior in specific ways. This paper explores the applicability of usage-based models in cases of language shift, in which immersion in language can no longer be assumed. Section 2 begins with an introduction to language shift, and presents evidence from two kinds of linguistic data in two kinds of language-shift situations: variability in the pronunciation of laryngealized resonants in St'át'imcets (2.1) and variability in the use of consonant gradation in Sointula Finnish (2.2). Section 3 discusses the implications of this evidence for usage-based models of language in terms of expanding our understanding of frequency effects (3.1) and of understanding how frequency interacts with other effects in shaping speech (3.2).

2. Language shift

Language shift refers to the situation in which one language is being replaced by another as the dominant language of a speech community. Two kinds of language shift are considered here. The first is indigenous language shift, in which a language spoken by an indigenous community is being replaced by the more widely spoken language in the area (generally the language that was introduced by European colonists). The second situation is heritage language shift, in which the language spoken by an immigrant community is being replaced by the dominant language in the area to which they immigrated. Both of these situations involve language attrition. In the latter case though, some variety of the language is still spoken elsewhere, in the country from which the immigrant population came. In the former case, the language is not spoken anywhere else.

2.1 Case study #1: St’át’imcets laryngealized resonants

St’át’imcets (Lillooet Salish) is a Northern Interior Salish language, comprised of two (possibly three) dialects: Upper and Lower (and Douglas, possibly a branch of Lower). It is spoken in 11 bands in the interior of British Columbia, in a roughly triangular area between Mount Currie (west), Pavilion (east), and Port Douglas (south). There are approximately 100 fluent speakers remaining, all over the age of 60 (www.uslces.org/uslces.html). These speakers are also fluent in English, the language of daily communication. As is typical of Pacific Northwest languages, St’át’imcets has a very rich consonant inventory, including a series of sounds called laryngealized resonants (hereafter abbreviated LR_s). These are complex sounds consisting of at least two gestures: an oral one and a laryngeal one. For example laryngealized [j^ʔ] involves tongue body raising to produce [j] as well as some degree of laryngeal constriction. These sounds vary along two dimensions: the *timing* between the oral and laryngeal gestures and the *realization* of the laryngeal gesture. In terms of timing, the laryngeal gesture can precede the oral gesture (pre-laryngealization), it can follow it (post-laryngealization), or it can occur sometime during the oral gesture (mid-laryngealization). In terms of realization, the laryngeal gesture varies from a complete closure to a subtle drop in pitch and/or amplitude. Cross-linguistic research has focused on timing, and has shown that pre-laryngealization is the most common pattern, at least for LR_s in onset position (Gordon & Ladefoged, 2001).

Phonetic studies of LR_s in St’át’imcets have shown that while general patterns in timing and realization can be identified, there is an enormous amount of variability both within and across speakers in how these sounds are produced. Bird et al. (2008) conducted a phonetic study of LR timing across three languages: Nuuchah-nulth (Southern Wakashan), Ntəʔkepmxcin (Northern Interior Salish) and St’át’imcets. They found that timing in St’át’imcets was correlated with syllabic position: laryngealization occurred as far away from the syllable nucleus as possible, such that LR_s were pre-laryngealized in onset position (VC^ʔ.RV), post-laryngealized in coda position (VR^ʔ.CV), and mid-laryngealized intervocalically (V.R^ʔRV). Examples in (1) below illustrate the general timing pattern in St’át’imcets.

- (1) Timing of laryngealization across syllabic positions²
- | | <i>Position</i> | <i>Example</i> | <i>English gloss</i> |
|----|-----------------|-----------------------|---------------------------------------|
| a. | Onset | lul. ^ʔ múł | ‘always jealous’ |
| b. | Coda | səm ^ʔ .xál | ‘to cut a hide into strips of things’ |
| c. | Intervocalic | xi.m ^ʔ mín | ‘to put something out of sight’ |

In a follow-up study, Bird (2008) focused on individual speaker variation and found that in fact, only one of the three speakers whose speech formed the basis of Bird et al.’s (2008) paper (S1 in Table 1) exhibited the overall pattern. The other two speakers exhibited simplified versions of this pattern: S2 had the same timing in onset and intervocalic positions (pre-laryngealization) and S1 had a single timing pattern across all three positions (post-laryngealization).

Table 1. Variability in LR timing across three fluent St’át’imcets speakers

LR position	S1	S2	S3
Onset	ʔR	ʔR	Rʔ
Coda	Rʔ	Rʔ	Rʔ
Intervocalic	RʔR	ʔR	Rʔ

¹ Superscript [ʔ] is used here to mark laryngealization, whether or not it involves a complete closure. Its position relative to the sonorant indicates timing: [j^ʔ] is an example of post-laryngealization (see for example Table 1 below).

² Note: transcriptions here are broad, except for the LR itself, which is specified for where the laryngeal gesture is timed relative to the oral one(s). Syllable boundaries are marked with periods.

In terms of realization, Bird (to appear) showed that the acoustic correlates of laryngealization were generally stronger in post-stress position than in pre-stress position: post-stress position was associated with more instances of complete glottal closure, greater pitch and amplitude dips, and longer durations of laryngealization. As with timing though, there was extensive variability in the precise realization of LRs. For example, of the LR tokens pronounced with some degree of laryngealization, 85% were produced with a complete closure for S1, 66% for S2 and 74% for S3. These percentages indicate that there was substantial variation both within speakers (no speaker produced a complete closure 100% of the time) and across speakers (no two speakers produced complete closures with equal frequency).

The variability found in the timing and realization of St'át'imcets LRs has two potential sources. First, it is possible that LRs are simply underspecified in terms of laryngealization, in the sense of Steriade (1995). This makes sense from a functional perspective: St'át'imcets contrasts plain vs. laryngealized resonants, but there are no contrasts within the laryngealized set. In fact, no languages have been identified that systematically and productively contrast different implementations of laryngealization, for example, pre- vs. post-laryngealization.³ It is plausible that the only requirement in terms of phonetic implementation is that LRs be perceptually different from their plain counterparts, with the details of how this is done left unspecified and therefore subject to variability.⁴ The second possible source of variability is language attrition. Dorian (1981) has noted that language attrition is associated with increased variability at all levels of linguistic structure, including the phonetic implementation of sounds.

It is likely that both underspecification and language attrition play a role in generating the observed variability in St'át'imcets LRs. However, since the focus of this paper is on language-shift situations, only the second source – language attrition – will be considered here. From the perspective of language attrition, we can think of within-speaker and cross-speaker variability separately, as resulting from two different processes: within-speaker variability arises as a result of a breakdown of the mechanisms that normally limit variability. Cross-speaker variability emerges as the result of the rise of idiosyncratic pronunciation patterns. These two processes will be explained in Section 3.1 below, as part of the discussion of the implications of the St'át'imcets data for usage-based models of language.

2.2 Case study #2: Sointula Finnish consonant gradation

Sointula Finnish is an example of a heritage language that is being lost. Sointula is a small community on Malcolm Island, off the North East coast of Vancouver Island, British Columbia. Originally settled in 1901 as a utopian society by a group of Finns dissatisfied with the state of affairs in their native Finland, it now has approximately 600 residents, of whom 80% have Finnish roots. Approximately fifty Finnish speakers remain in Sointula. Saarinen (2009) conducted an extensive study of language attrition among six Sointula Finnish speakers, focusing on consonant gradation. This morpho-phonological process affects the stem consonants /p t k/ and is triggered by adding a gradation-triggering suffix to the stem. Finnish grammarians refer to four types of consonant gradation (Hakulinen et al., 2004):⁵ *Direct Quantitative* refers to gradation in which a geminate consonant in the nominative singular stem (assumed to be the basis for

³ It has been argued that Hupa contrasts pre- and post-laryngealization in a restricted set of forms. See Golla (1977) and Gordon (1996) for details.

⁴ For further support for the idea of underspecified laryngealized resonants, see Kingston (1985) on timing of laryngeal gestures in obstruents vs. resonants.

⁵ The standard description is morphologically-based: it assumes that the nominative singular stem is the 'basic' stem, from which other morphologically-related stems are derived. Perhaps a simpler way of describing consonant gradation, from a purely phonological perspective, is to say that 'strong' forms (geminate; plosives) occur when the suffixed stem ends in an open syllable; 'weak' forms (singletons; fricatives and sonorants) occur when the suffixed stem ends in a closed syllable. This phonological description allows us to do away with the notions of *indirect* vs. *direct* gradation.

gradation) becomes a singleton in another stem form, e.g. $kk \rightarrow k$ (1)a below. *Direct Qualitative* refer to gradation in which a singleton consonant in the nominative singular stem surfaces as a lenited version of this consonant in another stem form, e.g. $k \rightarrow v$ in (1)b below. *Indirect Quantitative* refers to cases in which a singleton consonant in the nominative singular stem surfaces as a geminate in another stem form, e.g. $k \rightarrow kk$ in (1)c below. Finally, *Indirect Qualitative* refers to cases in which a singleton consonant in the nominate singular stem surfaces as a fortified version of this consonant in another stem for, e.g. $\eta \rightarrow k$ in (1)d below.

(2) Finnish consonant gradation

Type	Alternation	English gloss
a. Direct Quantitative	kuk. ka ~ ku. ka -n	‘flower’: nom.sg. ~ gen. sg.
b. Direct Qualitative	pu. ku ~ pu. vu -t	‘dress’: nom. sg. ~ nom. plur.
c. Indirect Quantitative	ri. kas ~ rik. ka -i.-ta	‘rich’: nom.sg. ~ part. plur.
d. Indirect Qualitative	rej. η as ~ rej. ka -i.-siin	‘tire’: nom.sg. ~ ill. plur.

As (2) below illustrates, the exact result of consonant gradation varies in highly complex and often unpredictable ways. Leiwo (1984) notes that consonant gradation is acquired relatively late by Finnish children, a further indication of its complexity.

(3) Diabolical ‘k’ (Hakulinen et al., 2004)

Alternation	English gloss
palkata ~ palk ka an	‘hire’: inf. ~ pres.1sg.
puku ~ puvut	‘dress’: nom.sg ~ nom.plur.
kulkea ~ kuljen	‘wander’: inf. ~ pres.1sg.
auri η ko ~ auri η gon	‘sun’: nom.sg. ~ nom.plur.
leka ~ lekat	‘sledgehammer’: nom.sg. ~ nom.plur.
reikä ~ reiät	‘hole’: nom.sg. ~ nom.plur.

Because of its complexity, consonant gradation is an ideal candidate for exploring the effect of language attrition on morphological structure. Saarinen (2009) focused on the use of direct quantitative and qualitative gradation in six English-dominant speakers of Sointula Finnish, across three generations: G2 (children of two Finland-born parents), G3 (children of two Finnish parents born in Canada), and G2.5 (children of one Finland-born parent and one Finnish parent born in Canada). Saarinen was primarily interested in the role of lexical frequency in predicting accuracy of consonant gradation, based on Bybee (2001). Her prediction was that high-frequency suffixed forms would be accessed as wholes and therefore would exhibit accurate consonant gradation. In contrast, low-frequency suffixed forms would be accessed through parts and consequently would *not* exhibit accurate consonant gradation. Her reasoning was that, because the ungraded nominative singular stems are the most frequent in the language, these would be selected instead of the appropriate graded stems in composing the suffixed forms; this would lead to incorrect (missing) consonant gradation.⁶

What Saarinen found was that consonant gradation was indeed being lost: G2 used it more accurately than G2.5, who used it more accurately than G3. Furthermore, qualitative gradation was affected more than quantitative gradation (likely because qualitative gradation is less predictable than quantitative gradation). However, the predicted frequency effects were not found: high- and low-frequency forms did not lead to significantly different levels of accuracy in consonant gradation. Instead, the semantic salience of the gradation-triggering suffix had a significant effect: suffixes with an equivalent form in English (the speakers’ dominant language) were correctly used, and associated with correct gradation. For example, suffixes corresponding to English prepositions (e.g. ‘in’) were used correctly, and the associated stems were correctly

⁶ For a detailed description of Saarinen’s methodology, the reader is referred to her (2009) MA thesis.

gradated. Suffixes without an equivalent English form were omitted and as a result the forms used were incorrectly gradated. For example, the object marker in Finnish is highly frequent, but it does not have a corresponding form in English since English does not mark syntactic roles morphologically. This suffix was often omitted and as a result consonant gradation was also omitted.

Saarinen's (2009) experimental design relied on frequency counts taken from the *Nykysuomen taajuussanasto* Finnish spoken language corpus.⁷ What her results showed was that these frequency counts were not relevant in Sointula Finnish: all words were infrequent for the speakers, and consequently expected differences in behaviour between what should have been high- vs. low-frequency forms (based on corpus data) could not be observed. On the other hand, the semantic salience of the target suffix acted as an excellent predictor of gradation accuracy, more specifically whether or not the suffix corresponded to a morpheme used in English, the dominant language of Saarinen's participants. The implications of these findings for usage-based models of language are discussed in more detail in section 3.2 below.

3. Discussion

Based on the patterns found in St'át'imcets and Sointula Finnish, what can we say about the applicability of usage-based models of language to language-shift situations? In this section we take a closer look at the factors at play in accounting for the data presented above, focusing on their implications for usage-based models of language. Section 3.1, based on the St'át'imcets findings, focuses on expanding our understanding of the ways in which frequency can affect phonetic variability. Section 3.2, based on the Sointula Finnish findings, addresses the question of how frequency interacts with other factors in affecting lexical access.

3.1 Expanding our notion of 'frequency effects'

Phonetic variability is one area of linguistic research in which usage-based models have figured prominently. Indeed, frequency is at the heart of usage-based explanations for phonetic variability. For example, it is argued that high-frequency words are more likely than low-frequency words to be reduced, which explains why we get ['mɛm.ɪ] instead of ['mɛm.əɪ] for *memory*, but we do not get *['mæm.ɪ] instead of ['mæm.əɪ] for *mammary* (Bybee, 2001). The following discussion focuses on Exemplar Dynamics, as outlined by Pierrehumbert (2001, 2002 and 2003), to illustrate the implications of the variability observed in St'át'imcets LRs for usage-based models of language. Two distinct frequency effects are considered, which have quite different effects on variability: a low-frequency effect creates within-speaker variability and a high-frequency effect creates cross-speaker variability.

According to Pierrehumbert (2003), our mental representations of sounds consist of sets (termed *clouds*) of tokens (termed *exemplars*), which we acquire by perceiving the speech around us. Speech production proceeds in the steps summarized in (4) below.

- (4) Steps assumed in speech production within Exemplar Dynamics (Pierrehumbert, 2003)
 - a. Select a category to be produced
 - b. Take a (random) selection from the exemplar cloud for that category
 - c. Entrenchment: average *n* neighbours of that selection to get a production target
 - d. Produce the target with some random error

⁷ *Nykysuomen taajuussanasto* [Modern Finnish Lexicon]. Aineistopalvelu Kaino. Kotimaisten kielten tutkimuskeskus [Research Institute for the Languages in Finland]. Retrieved from <http://kaino.kotus.fi/sanat/taajuuslista/parole.php> (2008).

Crucial to the discussion here is what is called *entrenchment*, the averaging process that occurs before a target speech sound is produced. Because it is an averaging process, its effect is normally to limit variability in the speech output. As Pierrehumbert suggests below, entrenchment requires ‘extensive experience’; the prediction then is that in language-shift situations in which extensive experience is no longer possible, the entrenchment mechanism breaks down, leading to rampant variability in speech production:

“... without very extensive experience with the dialect, errors in establishing the label set and effects of undersampling would combine to predict various kinds of over- and under-generalization in phonetic outcomes” (Pierrehumbert, 2002: 114).

Figure 1 illustrates the breakdown of the entrenchment effect. On the left is a non-shift situation: exemplar clouds are well populated. As a result, averaging occurs over a large number of exemplars and the token produced is an accurate representation of its category. On the right is a language-shift situation: the exemplar cloud is poorly populated. As a result, averaging occurs over a very small number of exemplars⁸ – in this case a single one – and the token produced is not likely to be very representative of its category.

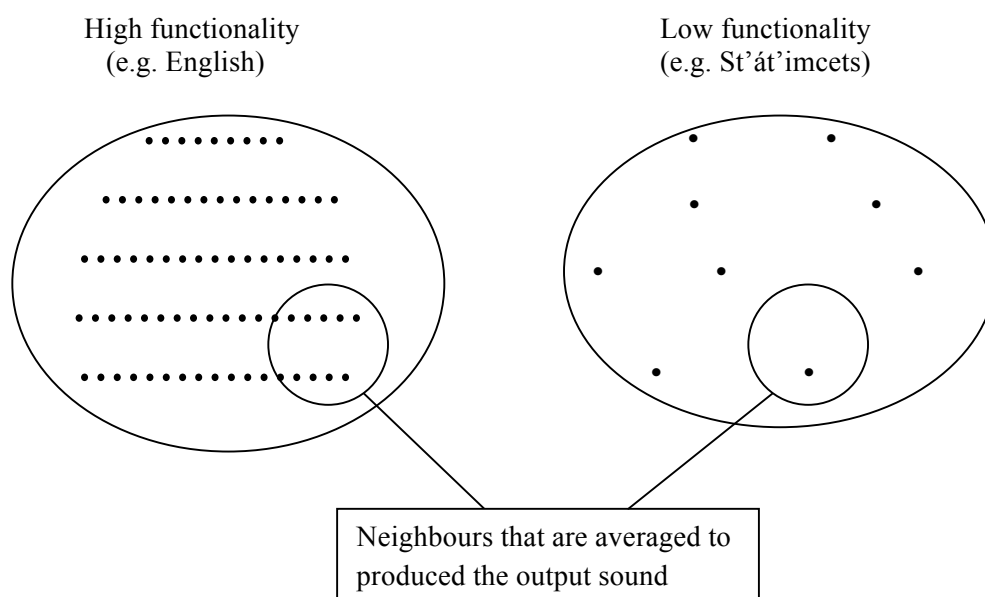


Figure 1: Loss of functionality and resulting increase in phonetic variability. On the left is a high-functionality (non-attribution) situation, e.g. English; on the right is a low-functionality situation, e.g. St’át’imcets. Figure reproduced from Bird (2008).

The effect illustrated in Figure 1 is a *low frequency* effect: language attrition results in low frequency for all speech sounds, which leads to lack of entrenchment and increased *within-speaker* variability: individual speakers are highly variable (i.e. inconsistent) in their pronunciations of LRs from one instance to the next because of limited exposure and use of these sounds. What is interesting here is that frequency cannot be used in St’át’imcets to categorize forms (individual sounds or bigger units) into frequency classes (e.g. high- vs. low-frequency, as in *memory* vs. *mammary* mentioned above). Rather, all sounds and words are effectively infrequent. In this sense, frequency is no longer a useful predictor of how sounds will pattern within the language – as it is in healthier language situations (see Bybee, 2001). This seems typical of language-shift situations; as we shall see in Section 3.2 below, the same situation is true for Sointula Finnish.

⁸ Assuming it is done over a fixed space as in Hintzman (1986) and Goldinger (1996).

Another frequency effect is at play in the St'át'imcets data, one that creates *cross-speaker* variability and that can be characterized as a *high frequency* effect. In situations of advanced language-shift, as is the case with St'át'imcets, the most common uses of language are a) ceremonial speech, in which the speaker does not interact with anyone else and b) one-on-one or small group conversations, in which the speaker contributes a relatively high proportion of the speech used (approximately half in a conversation with one other speaker, or one third in a conversation with two other speakers). The result is that the voice that speakers hear most frequently during language use is often their own. Within *Exemplar Dynamics*, because exemplar clouds are populated by exemplars perceived during language use, speakers' clouds will include their own pronunciations in relatively high frequencies; this leads to the reinforcement of idiosyncratic pronunciations, i.e. the emergence of systematic cross-speaker variability. Take for example S3 in Table 1 above, who only uses post-laryngealization. Initially, her preference for post-laryngealization was likely not consistent or intentional in anyway. However, once she started producing post-laryngealized LRs, this timing pattern quickly became the most common in her exemplar clouds (a snowball effect), and as a result she became more and more likely to produce post-laryngealized LRs over time, until her pronunciation did become entirely systematic and different from S1 and S2 (also in Table 1).

Summarizing, the St'át'imcets case shows us that language attrition leads to two separate frequency-related effects: 1) within-speaker variability results from the low frequency of all tokens (i.e. poorly populated exemplar clouds); 2) cross-speaker variability results from the high frequency of exemplars within individual speakers' clouds from their own speech. Expanding our notion of frequency to encompass these fairly distinct frequency effects, which are both typical of language-shift situations, will help us refine our understanding of the role of frequency in explaining the linguistic patterns observed across a wide range of linguistic contexts.

3.2 The interaction between frequency and other effects on lexical access

The previous discussion focused on phonetic variability. Another area of linguistic research in which usage-based models have figured prominently is lexical access. Recall that in Saarinen's (2009) work on Sointula Finnish, the prediction (based on Bybee, 2001) was that high-frequency forms would be accessed as wholes and therefore would be correctly gradated, whereas low-frequency forms would be accessed as parts and would be incorrectly gradated as a result of selecting the wrong parts. Contrary to the prediction, frequency was not found to play any role at all in gradation accuracy. Saarinen attributed her findings to language attrition: in Sointula Finnish, because the language is spoken so rarely, all forms are effectively low-frequency from the perspective of the speakers. This situation is analogous to the one discussed above with respect to LRs in St'át'imcets, but in the realm of lexical access rather than speech output.

Because all forms are low-frequency, corpus-based frequency counts cannot be used as a way of predicting how lexical access occurs in Sointula Finnish. Instead, the effect we see is one of dominant-language transfer: suffixes that are semantically salient to English-dominant Finnish speakers are retained and associated with correct gradation. In contrast, suffixes which are not salient are not retained, and gradation is lost when the gradation triggering suffixes are lost.

While usage-based models have tended to focus on frequency as the primary source of observed phenomena relating to lexical access, a number of researchers have argued that there is more to lexical access than frequency. Hay (2001), Hay & Baayen (2005) and Järvisikivi et al. (2006) have discussed lexical access as a function of suffix salience. They have argued that suffix salience (and consequently whether suffixes can be accessed independently from the stems to which they attach) is a function of a combination of factors, a sampling of which is provided in (5) below.

(5) Effects on suffix salience (Hay, 2001; Hay & Baayen, 2005; Järvisikivi et al., 2006)

- a. *Phonotactics*: a clear boundary between the stem and the suffix(es) makes the suffix(es) more salient
- b. *Suffix length*: longer suffixes (orthographically) are more salient

- c. *Allomorphy*: lack of allomorphs makes suffixes more salient
- d. *Homonymy*: lack of homonyms makes suffixes more salient

Sointula Finnish adds evidence in support of Hay and her colleagues, providing a strong argument for the idea that there is more to lexical access than simply frequency, particularly in language shift situations. In this case, with the effect of lexical frequency no longer relevant (because all forms are collapsed into a single low-frequency category), semantic salience has taken over as the best predictor of lexical access.

3.3 Summary

Usage-based models have emphasized the role of frequency effects on shaping speech patterns. The St'át'imcets data show us that the notion 'frequency' can be elaborated: different kinds of frequency effects are at play in language shift situations than in non-shift situations: low frequency of use leads to a lack of entrenchment and within-speaker variability in production; high frequency of speakers' own pronunciations within their exemplar clouds leads to the divergence of idiosyncratic pronunciations and to systematic cross-speaker variability. The Sointula Finnish data show us that frequency should not be viewed universally as the single, most prominent source of linguistic variability, as tends to be done in usage-based models. Many other factors contribute to shaping language as it occurs in speech communities, and this is particularly true in language-shift situations: in Sointula Finnish, the single most important effect on consonant gradation use is semantic salience, from the perspective of English-dominant Finnish speakers.

4. Conclusion

This paper has presented two case studies, illustrating the effects of language attrition in two kinds of linguistic data from and two kinds of language-shift situations. What these case studies show is that usage-based models of language can account for linguistic patterns found in situations of limited language use as long as they include the flexibility to allow for 1) an expansion of what we mean by 'frequency effects' and 2) a way of including other effects, which can take over when frequency-related effects are no longer relevant. This flexibility is indeed a necessary component of any model of speech production (and perception), if it is to reflect the reality of language use in a broad range of interlocutory interactions.

References

- Bird, S., Caldecott, M., Campbell, F., Gick, B., & Shaw P. (2008). Oral-laryngeal timing in glottalized resonants. *Journal of Phonetics* 36:492-507.
- Bird, S. (to appear). The nature of laryngealization in St'át'imcets laryngealized resonants. *International Journal of American Linguistics*.
- Bybee, J. (2001) *Phonology and language use*. Cambridge, UK: Cambridge University Press.
- Dorian, N. (1981). *Language death: The life cycle of a Scottish Gaelic dialect*. Philadelphia, PA: University of Pennsylvania Press.
- Golla, V. (1977). A note on Hupa verb stems, *International Journal of American Linguistics*, 43(4):355-358.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22:1166-83.
- Gordon, M. (1996). The phonetic structures of Hupa. *UCLA Working Papers in Phonetics*, 93:164-187.
- Gordon, M. & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics* 34 (1), 49-72.
- Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics*, 39(6):1041-1070.

- Hay, J., & Baayen, R.H. (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Science*, 9(7):342-348.
- Hintzman, D.L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review*, 93:328-338.
- Järvikivi, J., Bertram, R., & Niemi, J. (2006). Affixal salience and the processing of derivational morphology: the role of suffix allomorphy. *Language and cognitive Processes*, 21(4):394-431.
- Kingston, J. (1985) *The phonetics and phonology of the timing of oral and glottal events*, PhD dissertation, University of California.
- Pierrehumbert, J. (2001). Exemplar Dynamics: Word frequency lenition and contrast. In Bybee, J., & Hopper, P. (eds), *Frequency and the emergence of linguistic structure*. Amsterdam, Holland: John Benjamins Publishing Company, 137-157.
- Pierrehumbert, Janet. (2002). Word-specific phonetics. In Gussenhoven, C., & Warner, N. (eds), *Laboratory Phonology VII*. Berlin: Mouton de Gruyter, p. 101-139.
- Pierrehumbert, Janet. (2003). Exemplar Theory. Paper read at the 77th meeting of the Linguistic Society of America, Atlanta, GA.
- Saarienen, P. (2009). *The Finnish language in post-utopian Sointula: The effects of frequency on consonant gradation*. MA thesis, University of Victoria.
- Steriade, D. (1995). Underspecification and markedness. In Goldsmith, J.A. (ed) *The handbook of phonological theory*. Cambridge, MA: Blackwell Publishers. Pp. 114-174.

Sonya Bird
sbird@uvic.ca