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**

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 environmental 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 phylogenic 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 at 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 priors[1] 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):

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

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 & Seindeberg, 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 magnificient 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., Nespor, M. & Bonatti, L. L. (2006). *The "Soul" of language does not use statistics: Reflections on Vowels and Consonants.* Cortex, 42, 846-54

Nespor, 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., Nespor, 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.

Reali, 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 Nespor, M. (2008) *The vowel tier constrains statistical computations over the consonantal tier*. Poster presented at AMLaP, September 2008, Cambridge, UK.

Shukla, M. & Nespor, M (in press) Rhythmic patterns cue word order In: L. Rochman, & N. Erteschik-Shir (Eds.) *The Sound Pattern of Syntax.* OUP: Oxford.

Shukla, M., Nespor, 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

Swingley, 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)dimilarities. *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., Nespor, 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