Language Sciences 53 (2016) 58–74

Contents lists available at ScienceDirect

Language Sciences

journal homepage: www.elsevier.com/locate/langsci

How spoken languages work in the absence of an inventory of discrete units

Michael Ramscar, Robert F. Port*

[a] Eberhard Karls Universität, Tübingen, Germany
[b] Indiana University, Bloomington, IN, United States

ARTICLE INFO

Article history:
Available online 26 September 2015

Keywords:
Language
Speech perception
Language understanding
Linguistics
Learning theory

ABSTRACT

Historically, linguists and psychologists have generally assumed that language is a combinatoric process, thereby taking the idea that language users have access to inventories of discrete, combinable units (phonemes, morphemes, words, etc.) for granted, despite the fact that these units have tended to resist formal definitions. We propose a new approach to language understanding based on the psychological mechanisms that underpin context-sensitive processing. This new method is surprisingly simple, in large part because it embraces a view of learning that has been developed from studies of animal behavior and neuroscience. From this perspective, learning is seen as a systematic, discriminative process that seeks to reduce a learner's uncertainty in making moment-to-moment predictions. We suggest that language processing employs all the information available to the listener at any given moment to predict what will happen in the next moment, in the next couple of sentences, etc. This approach does not rely on any of the ambiguous traditional linguistic units because continuous-time processing simply acts to reduce a hearer's uncertainty about an actual message in relation to possible messages, rather than building up an interpretation out of elemental components. From this perspective, the conventional units of language – phonemes, morphemes, words – can be seen as idealizations of patterns that evolved for communicative efficiency that can serve the purposes of orthographic (and linguistic) description, rather than psychologically 'real' elements that are essential to language processing.

1. Introduction

Many linguists and laymen believe that natural languages are made up of discrete units of various kinds: (1) *words,* which can be easier or harder to identify (e.g., *dog, table*, *dunno*); (2) *phonemes,* roughly the acoustic/psychological version of letters; and (3) *sentences*. Another somewhat technical unit is (4) the *morpheme*, a fraction of speech that has a consistent meaning and cannot be divided into smaller such pieces (e.g., *tree-house, rain-bow-s, walk-ing* = 7 morphemes total). Linguists generally consider these four units to be widespread across human languages, because in many languages utterances appear to be composed out of sequences of words, words appear to be composed out of one or more morphemes and morphemes in turn appear to be composed out of phonemes. These units are often considered to be universal and to be essential

* Corresponding author.
E-mail address: port@indiana.edu (R.F. Port).

http://dx.doi.org/10.1016/j.langsci.2015.08.002
0388-0001/

components of language. The semantic interpretations of utterances are thought to be constructed along similar lines, by combining the meanings of the morphemes that comprise words that in turn may comprise a sentence. This view of language is usually called *compositionality*, and virtually all of modern linguistic theory is predicated on this principle.

The twin ideas of units and combinatorics have powerful intuitive appeal. However, most writing systems overlook many of the important contrasts that are realized in speech (such as intonation, stress or the different acoustics of the final sibilant in *pots* and *dogs)*, and of course, people who ponder the formal nature of language invariably have a high level of literacy, which raises the possibility that lifelong reading practice biases the theoretical intuitions of researchers. However, when critical attention is focused on the units of spoken language, they cannot be specified with sufficient clarity to play the central role that these linguistic theories attribute to them.

In this paper, we focus on the problem of speech perception and comprehension by adults leaving the important issue of the production of speech to another day. We will describe some of the many empirical and theoretical reasons that cast doubt on the idea that human speech comprehension relies on discrete phonemes, words, sentences, or any other units. We then sketch an alternative account of the workings of language based on learning theory. Critically, our account does not depend on language users being in possession of inventories of discrete components. Rather, from a learning perspective, language and linguistic knowledge can be seen to be fundamentally continuous in time and systematic in nature. This account treats languages as cultural systems that have been shaped by the constraints imposed by learning and communication. Indeed, the dynamic nature of language systems at both the communal and individual levels help explain why it is that theories of language based on units have inevitably proven so unsatisfactory when it comes to accounting for and explaining the facts of linguistic communication.

Why units don't work. While it cannot be denied that each language comprises a system of contrasts that yield patterns that do resemble all four of the units we listed above, research has shown that none of these units can be given clear definitions, and none of them can be consistently applied to real speech. None can play the role of discrete symbol tokens analogous to letters as required by formal theories of language (Chomsky and Miller, 1963; Chomsky and Halle, 1968). On the other hand, it is also clear that many cultures have developed discrete *orthographic systems*. However, although letters, words and sentences (or their equivalents) can be identified in most writing systems, attempts to specify corresponding units in speech that are reliable have foundered.

In the view we will sketch in this paper, language perception proceeds by means of a continuous-time process involving a learned system: a listener's current knowledge of a language and its relationship to the world, which comprises a series of relationships between semantic distinctions and acoustic phonetic information. Learning is a discriminative, predictive process (Ramscar et al., 2010a,b, for a review) that reduces a learner's uncertainty about the world. Comprehension involves a continuous series of predictions about upcoming semantic and acoustic patterns that reduce a listener's uncertainty about an intended meaning by a process of elimination. A listener's understanding of the intended semantic message of the speaker is gradually revealed as alternative messages are successively rejected. The important thing is the continuous interaction between semantic and acoustic phonetic aspects of the language system.

To give a concrete example, suppose someone says:

A. *Wanna nother drink?* (Do you want another drink?)
B. *What's 'at?* (What is that you said)?)
A. (Picks up a soda can.)
B. *Yeah, I'd love one*.

The context is presumably an occasion where A, the hostess, has offered a drink to B, a guest. B, saying *What's that?* (with stress on *what,* not on *that)* reveals that he is uncertain what she said, so instead of repeating herself, she raises a can of drink into B's visual field, thereby reducing his uncertainty about both what it was that she said and her intended meaning.

Let's look at B's response in more detail. B sees the soda can and this reduces his uncertainty about what A had just said sufficiently to allow him to say that he would actually like another drink. The phonetic pattern [jæ:] by B indicates acceptance of the offer and [aɪdlʌv …] (*I would love …*) accepts politely by implying that the host is still free to retract the offer. A third person overhearing this utterance can use the [jæ:] to partially predict B's *I'd* (rejecting *No thank you*) and using *I'd* to partially predict *love* (and reject *rather not…* etc.*)* and *love* to predict *another (*and reject *a couple of 'em*, as well as *Fresca* and *I'd love to but …* etc.). The point is that both acoustic phonetic information (acoustic and visual) and semantic (and social) context work together from moment to moment to reduce a listener's uncertainty about a speaker's intended message.

Although real-time predictions do not generally lead to certainty before the word begins, simply narrowing the range of possibilities is very valuable to a listener. Since the phonetic and semantic sides of the language system work continuously together, people in some situations may be quite unaware of complete interruptions of the speech signal. If someone scrapes a chair noisily across the floor to mask the acoustics of the entire word *love*, listeners will normally believe they heard the word *love* because their language system fully predicts it based on the previous (and following) acoustic and semantic context (Warren and Obusek, 1971). From the perspective we describe here, "speech perception" can be seen as a similar kind of continuous process in which the phonetic and semantic aspects of speech are inseparable. Speech understanding is the product of continuous-time learning that constructs a system of phonetic and semantic contrasts that enable a speaker's intended message (which is, itself, situated within an acoustic phonetic and semantic system) to be discriminated from other possible messages a speaker might have uttered. We put quotes around *speech perception* above because the view of speech

perception we describe here sees this as a systematic process in which a listener uses prediction and the integration of subsequent information in order to reduce uncertainty about a speaker's message, as opposed to more traditional accounts in which a listener first perceives the acoustic phonetic form of a message, then accesses the lexicon and then decodes a meaning from it.

We turn now to criticisms of the various linguistic unit inventories that are widely believed to be essential components of language.

2. Phonemes

The idea that all the words of any given language could be spelled with a limited set of sound categories called *phonemes*, appropriate for each language is over a hundred years old. It can be dated to the late 19th century, and developed roughly simultaneously with the development of the International Phonetic Association (IPA) *phonetic alphabet*. Phonemes were an attempt to capture an undeniable tabular structure apparent in the words of many languages, such that the orthographic representations of speech contrasts can be lined up so that roughly the same set of vowels occur in a large number of consonant contexts:

beat, bit, bait, bet, bat, bot
mead, mid, made, med, mad, mod

and consonants can also be organized by similar tables (initial consonants on the left, syllable-final ones on the right):

mad	*bad*	*pad*	*fad*	*lamb*	*lab*	*lap*	*laugh*
nod	*Dodd*	*Todd*	*sod*	*can*	*cad*	*cat*	*Cass*
—	*god*	*cod*	—	*lung*	*lug*	*luck*	—

Many 20th century linguists (e.g., Jakobson et al., 1952) took the success of phoneme-based writing systems to infer that phonemes are all one needs to specify words – that, in fact, phonemes really are formal tokens so that all one needs to produce or perceive a word in any language is the specification of its phonemes or the "distinctive features" that are used to differentiate the phonemes from each other. They were taken to be formal tokens, just like letters as used in mathematical grammars.

However, treating phonemes as discrete tokens obscures the fact that the vowel prior to the/d/ in *mad* (or the/b/ in *lab*) is longer than that in *mat* (or in *lap*). These different variants of the "same" phoneme, called *allophones*, make it clear that the phoneme itself, as an abstraction, cannot actually be the basic unit of speech. Details of timing are essential. A further level of granularity was postulated below the phoneme, the *phone,* which represents a supposedly more concrete sound type than the phoneme – yet still consisting of timeless, serially–ordered 'segments'. Chomsky and Halle (1968) and essentially all subsequent formal phonologists have assumed (without serious defense) that phonetic features for the phones comprise a universal alphabet of "speech sounds" underlying all languages – "the phonetic capabilities of man" (p. 295) as Chomsky and Halle (1968) put it. Many linguists behave as though all they need to know about speech or phonetics is what is provided by the phonetic distinctive features.

The problem – and the reason for the quotes around "speech sounds" – is that it has been understood since the 1940s (Joos, 1948) that physical speech signals are continuous events (or 'quasicontinuous' as Fant, 1959 put it), overlapping in time such that it is impossible to identify acoustic slices (that is, static spectra integrated over a short time window) corresponding to each phone or phoneme. Indeed, because it is has proven impossible to objectively count phones or phonemes on speech acoustic displays, counting "speech sounds" has traditionally relied on having some person listen to a speech sample and count the letters they would use to write it down. Automatic speech recognition (ASR) has had some success at machine transcription of speech. However the more successful current systems usually employ language models that use probabilistic representations of the likelihoods with which linguistic events will happen in context (i.e., Shah et al. 2014; Kim et al., 2015). These ASR systems more closely resemble the approach we describe here than they do traditional ideas regarding speech perception. However, it remains the case that some linguists and phoneticians continue to speak of speech sounds as though there really were letter-sized units in the speech signal that can be counted.

Phones, Phonemes and the Orthographic Alphabet. Historically, the IPA alphabet was constructed by starting with the orthographic alphabet of Latin as used in western European languages and then adding letters inspired by other alphabets in order to formalize a system in which a single letter consistently represents a single "speech sound." The obviously continuous nature of vowels was dealt with by adding diacritic arrows (e.g., $>$, $<$, $\wedge$, $\vee$) to suggest backer, fronter, higher or lower versions of any of the set of vowel symbols. Of course, phoneticians generally do not assume that the phonetic alphabet models all possible speech sounds. The IPA alphabet is treated as an expansible tool for practical use in describing speech sounds. Of course, phones share many of the difficulties of phonemes.

On the basis of this history, it seems that many language scientists assume that phones and phonemes are psychologically and linguistically real, and that they provide a discrete minimal spelling system for storing, producing and perceiving the words or morphemes of individual languages. Some even assume they are derived from a universal, that is, innate, inventory of segmental phone types (Chomsky and Halle, 1968). However, researchers have uncovered many reasons to doubt that any of these assumptions are valid (Faber, 1992; Harris, 1998; Öhman (2002); Linell, 2005; Port and Leary, 2005; Port, 2010).

Why phonemes cannot provide 'psychological spellings' of words. The most obvious and fundamental limitation of abstract phonological segments is that, despite their descriptive usefulness and intuitive appeal, they fail to capture a great deal of critical information relevant to the perception and production of language. Empirically, it is clear that words are stored, produced and perceived with far more detail than the idea of phonemic or phonetic letters indicates:

- **Speaker Imitation.** One kind of evidence comes just from our experience as attentive listeners to speech. For example, if RFP says the words *bat* and *bad* in his native American dialect, consistent small differences in quality and duration between the two [æ] vowels will be noticeable. If he imitates the distinctive pronunciation that speakers from New York City typically use for *bad*, many readers would likely recognize that he is imitating NYC speech, discriminating it from his normal speech. The point is that most speakers can detect small, sub-phonemic temporal and spectral differences in word pronunciation due to dialect variation and even imitate them in many cases. It follows that the details that allow people to discriminate and imitate these differences must be stored in memory.
- **Minimal Pairs.** The notion of "minimal pairs", that is, word pairs (or triples or more) that are words that sound different, is often insufficient to specify phoneme identity across words. Thus, the clear difference (at least to English speakers) between *bead* vs. *bid* ([bid] vs. [bɪd] and *seed* vs. *Sid* ([sid] vs. [sɪd]) establish that/i/and/ɪ/should be considered different phonemes. However, there is a problem. This contrast does not exist before an/r/(at least for speakers who pronounce/r/ after a vowel). So words like *beer, spear, hearing* are usually assumed to share the "same vowel", usually close to [i]. But there are no words with [-ɪr] that contrast with [ir]. Thus a contrast found elsewhere in the vocabulary is collapsed, or "neutralized," before a syllable-final/r/. In this case there is no definitive answer to the question of which phoneme should be used. *Beer* might be transcribed with either [i] or an [ɪ] because there is no contrast. This is a problem for the claim that words are always spelled from a set of phonemes. To deal with this issue, one might even postulate a new vowel phoneme that sounds much like/i/but which does not imply a contrast between [i] and [ɪ].
  This problem arises often in many languages. Another example from English is the loss of distinctiveness of "voicing" as found in/t/vs./d or/p/vs./b/when a stop occurs after/s/and before a vowel at the beginning of a syllable. Thus words like *spot*, *stop* or *Scot* are spelled with P, T and K, but there is no possibility of words with the phonetics of/b/,/d/or/g/following/s/ and before a vowel. If one pronounces these words slowly and listens, one will find that the P, T and K in such words actually resemble a/b/,/d/or/g/. If the [s] is deleted (or artificially separated) from a recording of the word *Scot,* the remainder sounds more like *got* than like *cot.* These and many other examples show that minimal pairs cannot always be relied upon to determine what the actual phonemes in a word are. Unfortunately, they produce indeterminacy about what "psychological spelling" should be for these words.
- **Recognition Memory Tasks.** Other evidence for the inadequacy of phones and phonemes comes from experiments analyzing auditory recognition memory, in which participants listen to lists of spoken words, and push a button after each word to indicate if the word is new or a repetition. Unsurprisingly, as the number of intervening items increases, accuracy decreases. However if several different voices are used to present the words, listeners are more accurate when repetitions are heard in the original voice than in a new voice (Palmeri et al., 1993; Lachs et al., 2003). This finding clearly contradicts that idea that words are stored in memory using a phonemic code, because purely phonemic representation would abstract away from properties specific to individual voices (Chomsky and Halle, 1968; International Phonetic Association, 1949).[1] Thus, these results suggest that speakers discriminate words using much richer information than phonemes.
- **Ubiquity of Important Timing Detail.** Orthographic alphabets comprise a sequence of symbols (letters, numbers or some other set of tokens), which permit no measure of continuous time. The only method by which time can be measured using an alphabet is a primitive one: to indicate greater length by additional symbols: [baada] should have a longer vowel than [bada] and [batta] a longer stop than [bata]. However, timing and rhythm in language is far more varied than phonemes themselves imply.

For example, in English many words are differentiated by the relative durations of consonant and vowel intervals (Klatt, 1976). Looking at pairs like *fuzzy* and *fussy* (or *rabid/rapid* or *bigger/bicker*), the vowel in *fuzzy* is longer than in *fussy* while the medial consonant duration is shorter – similarly for the other pairs (Klatt, 1976; Port, 1979). Timing detail is critical to the correct pronunciation of words in most languages (Klatt, 1976) and perturbations of timing make speech difficult to understand (Tajima et al., 1997). This helps explain why second-language speakers can be difficult to understand for native speakers.

These differences, along with many other factors that are consistently reflected by timing variations – such as a speaker's focus on important words (Connine et al., 1987; Lindblom, 1990), the phonological neighborhood density of words, or the predictability of words in context (Gahl, 2012) – are, of course, invariably left out of phonemic transcriptions despite their importance for listeners.

There is further evidence that subphonemic variation plays an important role in both speech production and understanding. For example, speakers produce nouns with longer mean durations when they are pronounced as singulars than as the stem of the corresponding plural (Baayen et al., 2003). Importantly, listeners are sensitive to the way that speakers

[1] One might propose the listeners adopted a strategy of storing some properties of the speaker's voice in addition to the "phonemic code". However, the results exhibited no difference at all between the two-voice results and the multivoice results – even when there were 20 voices. These results strongly suggest that speakers routinely maintain a great deal of phonetic detail whenever they hear speech.

discriminate between the vowels of the "same" word in its plural and singular form (Kemps et al., 2005). When the segment that linguists usually consider to be added in marking a Dutch noun as a plural is spliced out of a recording, native speakers do not perceive the noun to be a singular, but rather they tend to hear a distorted plural. As Kemps et al. (p. 441) observe: "The way words are written in languages such as Dutch and English suggests that they consist of stems and affixes that are strung together as beads on a string. Phonemic transcriptions convey the same impression. Our experiments show [that] … plurals are not just singulars with an additional suffix."

There is clear evidence that what listeners perceive in natural speech is heavily guided by their expectations (for review, see Cutler, 2012). Thus when the highly reduced forms of very frequent words are heard out of context, listeners are unable to recognize them. For example, when native Dutch speakers are asked to report exactly what they hear and then are presented with speech resembling [ɛik] (a reduced form of the Dutch word *eigenlijk;* "in fact") without context, they report hearing *eik*. However, when listeners heard the same audio clip presented in their full context, they hear *eigenlijk,* "in fact" (Ernestus et al., 2002).

Further, the more well-learned an acoustic/semantic contrast is, the more likely it is that it will be "heard" in a highly reduced form for which there is little or no acoustic evidence. Thus in phoneme-monitoring experiments, listeners will actually claim to perceive the presence of frequently encountered suffixes that are in fact partly and even completely missing in the reduced acoustic signals that they hear (Kemps et al., 2004).

In recent years, new research paradigms have blossomed by phonetician-linguists seeking to provide dynamical models of speech (Port et al., 1995; Browman and Goldstein, 1995; Gafos et al., 2012). These approaches, under the general heading of laboratory phonology, greatly enrich the description of speech beyond the segment-only descriptions of traditional linguistics and generative phonology. Instead of studying just phones and phonemes, these workers create models employing continuous time and make use of point attractor and oscillator dynamics to describe the simultaneous gestures of speech. This work makes use of greatly improved representations of speech acoustics as well as continuous descriptions of speech articulation, providing novel ways to understand speech. Certainly, this program provides further support for speech based on continuous time, non-segmental descriptions.

It would appear that no matter how convenient phones and phonemes may be as descriptions and no matter how intuitive they are to us, it is clear that whatever the representations language users may employ, they must be far richer than any version of phones or phonemes suggests. Only such rich versions of words could be psychologically or linguistically real. This in turn reveals how traditional ideas about language as composed from a small inventory of phonemes or distinctive features for word spelling have been unsupported through many decades of research.

3. Words and morphemes

If laymen are asked about "the units of language," their responses will likely focus on "words." Word units are highlighted in many orthographies where words are separated from each other by a blank (although in the orthographic systems of many Sinosphere languages, the question of word status is far less easily solved, Honorof and Feldman, 2006). From a linguistic perspective, words are often described as being composed of "morphemes," a term that supposedly refers to a minimum unit of speech with a consistent meaning. Thus, *banana* is a single word and a single morpheme but *bananas* is a single word with two morphemes (and *unreliable* is three: *un-rely-able*).

Formally, definitions of words and morphemes share many of the same difficulties. In a text, one can examine each letter and say what word it belongs to, and a space always separates two words. In speech, however, this is impossible for many reasons (Wray, 2013):

- **No audible separation.** There is nothing auditory that corresponds to the space on a page to separate fluently spoken words. (Potter et al., 1947; Joos, 1948; Fant, 1959).
- **Words can overlap in time** or even completely merge: in pronunciations of *What did you see?* [wədʒə] often merges part of *what* and all of *did* into the single [d] while the [ʒ] merges part of *did* and part of *you,* becoming *Wadja see?* when a writer wants to convey informal speech. (Note that the morphemes overlap here as well as the words.)
- **Orthographic conventions are inconsistent**: Many word boundaries seem completely arbitrary and inconsistent but survive simply due to tradition. It is interesting that expressions like *no one, of course, White House, on the other hand, even so, for that matter, for example, and so on, on your mind,* etc. are not spelled as single words, whereas expressions like *nothing, blackbird, heretofor, maybe* and *into* are. Despite orthographic tradition, it seems unlikely that *nobody* and *no one* really are qualitatively different kinds of psychological units.
- **Frequent collocations:** Many multiword sequences (e.g., *what do you think*) occur more frequently than seemingly common words (e.g., *learn, pet*) that might be thought of as the core vocabulary of English (Bannard and Matthews, 2008). Indeed, some sequences (e.g., *a cup of tea*) occur with a frequency almost equal to the frequency of their component words (e.g., *cup*, *tea*). Given these patterns of frequencies, and the fact the even very young children are sensitive to them over and above the frequencies of the words they comprise (Bannard and Matthews, 2008), it seems implausible that an efficient language processor would make the (orthographic) word the locus of processing, rather than whole phrases. As above, for many multiword sequences, the idea that they are stored as single units makes more sense than the idea that an orthographic word is a unit (see also Bolinger, 1975; Erman and Warren, 2000).

- **Agglutinated words.** More difficulties for definitions of words arise out of the agglutination found in languages like Swahili, Zulu and Turkish. In these cases, what is written and stressed as a single word has many possible prefixes and suffixes added to the word stem, making the number of potential "words" in the language soar. For example, a Swahili transitive verb can consist of up to 6 morphemes: First, a prefixed, monosyllabic pronoun for the subject of the verb (one of about 12 options), then an object pronoun (also a set of about 13), then the tense (one of about 5), then the verb stem followed optionally by one of at least 9 fully productive derivational suffix combinations. It then has one of 3 suffixes for "mood" (i.e., indicative, subjunctive or negative).[2] This yields over 21,000 possible inflectional variants of any verb. Should these all count as distinct words in Swahili, or should we count an individual set as just a single word? And what about more or less frequent inflectional variants (since it is known that many inflectional variants may never occur, Blevins, et al., 2015)? As with determining single and multiple word units in English, any decision made will tend to be arbitrary and poorly motivated.
- **Graded Degrees of Morpheme-ness.** Psychologically, the evidence for dividing words into morphemes seems to support a continuous graded property, not a binary distinction of morpheme vs. non-morpheme (Hay and Baayen, 2005). Priming experiments show that although one word can be used to facilitate the recognition of another under noise, these benefits vary continuously: *lead* strongly primes *leader*, but *dress* only weakly primes *dresser* even though both *leader* and *dresser* contain the "same" agentive *–er* suffix.

These observations show that, despite our strong intuitions that words, roughly as defined orthographically, are essential units for spoken language, the speech chunks that play the most critical roles can come in various sizes and may have little correspondence to the words of our orthography. For orthographic words, all that is required is a convention. For psychological words, the criteria seem impossible to identify.

So far we have seen that that none of the traditional discrete units of language that are thought to be the compositional parts of language – phones or phonemes, morphemes, and words – can be identified unambiguously enough to serve as the theoretical bedrock on which to build a successful theory of spoken language composition and interpretation. Surely, if these units were essential components of speech perception, they would not be so uncertainly identified. It would appear that the traditional and intuitively appealing idea of nested constituents from sound segments to words, relies on a faulty extrapolation from the orthography of many languages. While it is true that alphabetic written languages employ a fixed inventory of characters, it does not appear that linguistic utterances consist of hierarchically nested meaning-bearing constituents of progressively larger sizes.[3]

Moreover, orthographic coding is clearly context sensitive itself. For example, in reading English, the string *read* is read differently depending upon the grammatical role suggested by the local context. We know that the meanings of words change quite dramatically depending on context, and it is frequently not clear where boundaries between word or morpheme units should be placed. Moreover, speakers with different exposure to dialects and with different linguistic needs and experiences are certain to apply language differently.

So, it seems clear that for communication to succeed there must be sufficient common ground across speakers in their system of speech gestures and the patterns of meanings they wish to express. It is also clear that an accurate description of language could not insist on absolute identity between speakers' systems. It seems quite implausible that communication might rely on isomorphism between speakers' systems, because of (a) the wide variety of redundant yet different acoustic cues that hearers rely upon, (b) differences in linguistic exposure implying that many utterance fragments will have somewhat different meanings for different speakers and (c) the enormous and constantly growing size of, for example, the English vocabulary which guarantees that speakers will be familiar with differing samples of it (Ramscar et al., 2014).

For these reasons, we propose that the unit-like patterns in language are largely irrelevant to actual adult language use. Instead, language hearers employ the same continuous-time learning methods employed for other behaviors in order to learn how a vast number of speech cues (that would be considered context-sensitive in more traditional approaches) predict what messages (both phonetic and semantic) an interlocutor is attempting to express.

4. Toward a discriminative theory of language

For over a hundred years, linguists and psychologists have been trying to conceptualize language along the lines of the levels of the formal orthographic system (e.g., letters, words and sentences), as an inventory of elements that can be combined to serve the function of meaning transmission (Saussure, 1916). This effort is understandable, since it can seem as if morphemes "have" meanings,[4] especially when we use context-free short sentences of common words as examples to think about. If we listen to the spoken sentence "*The boy kicked the ball*," the meanings of *boy, kick* and *ball* can appear to be discrete

[2] An example of such a Swahili verb in a sentence would be *tu-li-m-l-ish-i-a ng'ombe* (we-Past-him-eat-Causer-Beneficiary-Indicative cows), "We fed (his) cows (for him)." (Port, 1981).

[3] We have ignored the sentence so far. Although we will not go into it, we consider the sentence to be one kind of linguistic construction that was regularized as a convention originally to make written language easier to understand. Of course, the convention of insisting that all speech ought to be produced in sentence form has been adopted as an ideal by most highly literate speakers.

[4] Dictionaries are largely predicated on the idea that words have context-independent meanings. However, the problems inherent in this are revealed with examples like *set*, where context has to do all the work: *set fire, set a table, set jello, the sun sets, chess set, set theory, set rules, get set,* etc.

and context-independent. Although tokens of sentences like "*The boy kicked the ball*" occur frequently, the overwhelming majority of sentence types are more akin to "*Her father threw her a ball*," or "*The monk kicked his habit*", which are ambiguous. This means that any intuitions we might have about context independent meanings are actually only applicable to a tiny fraction of the linguistic variation we encounter.

Lexical distributions. It is likely that just as orthography has had a distorting effect on theoretical intuitions, the peculiar statistical properties of language have also led linguists astray. Despite our conviction that words and morphemes play no essential functional role in spoken language comprehension, statistical analyses of language must rely on orthographic corpora for which counts of orthographic words are easily obtained. It has long been known that word and n-gram frequencies (an n-gram is a sequence of *n* contiguous items drawn from a text) are highly skewed (Fig. 1; Zipf, 1949; Baayen, 2001). Consider the relationship between word types (e.g., *dog*) and tokens (how often *dog* occurs). In English, a few words occur very frequently (e.g., *the*, *and*), such that half of the tokens in any large natural sample will be tokens of only 100 or so high-frequency types. The relative frequency of these types decreases rapidly (the most-frequent word may be twice as frequent as the second-most frequent), and frequency differences between types decrease as their relative frequency declines. This means that the other half of a large natural sample will be composed of ever-fewer tokens of a very large number of types, with ever-smaller frequency differences between them. Typically, around half of these types occur just once.

As well as being highly skewed, language distributions also exhibit bursts of particular words (Church and Gale, 1995; Katz, 1996). Discussions of word and n-gram frequency often speak of word-frequency as though it were a fixed property of a word, implying that texts can be treated as frequency-weighted random samples of the words of a language. But this is known to be incorrect. Statistically, all texts over-represent a few types of words while under-representing the majority of word types. Although the frequencies of the most common words in most texts will somewhat reflect their average frequency across many texts, most words have frequencies that are over-represented in a few texts but under-represented in most texts. Thus if we examine 100 texts at random, the rate at which common words like *and* and *the* occur will be fairly consistent. But when it comes to rarer words like *phoneme* or *n-gram* (or even *cow*), it's likely that we will not find either in any of our 100 random texts. And when we do find texts that contain *n-gram* or *phoneme* (such as this one), we will encounter these words far more often than we might expect based on their average rate of occurrence across all texts.

These statistical facts have important consequences for linguistic theory. The statistical facts of language actually mean that apparently "literal" sentences like "*The boy kicked the ball*" are highly unrepresentative of language as a whole. Only a small fraction of the types of sentences that are possible – or even attested – in any language will seem 'literal' in this way although there will still be very many literal sentence tokens. Indeed, the distribution of n-grams in languages inevitably means that linguistic intuitions based on considerations of "literal" sentences like "*The boy kicked the ball*" will mislead theorists about the nature of language because these sentences, while seemingly typical, actually represent only a tiny fraction of the variance among sentence types.

We suspect that it is a combination of these statistically unrealistic intuitions about literal meanings and the fact that linguistic study began with the examination of written texts that have led to the historical tendency to see linguistic communication as involving a process of encoding units of meaning into a set of lexical and grammatical forms and then transmitting and decoding these units. The challenge facing both language learners and theoretical linguists has been thought to be inductive: both the learner and the theorist have been viewed as facing the problem of inferring the correct inventory of form and meaning units along with the correct rules for combining and decomposing them. Yet, as we have sought to show here, even when it comes to the relatively less complex task of deciphering an inventory of form units, research indicates that this inductive challenge is not something that can be solved simply by proposing "underlying" or invisible or unlearnable components.

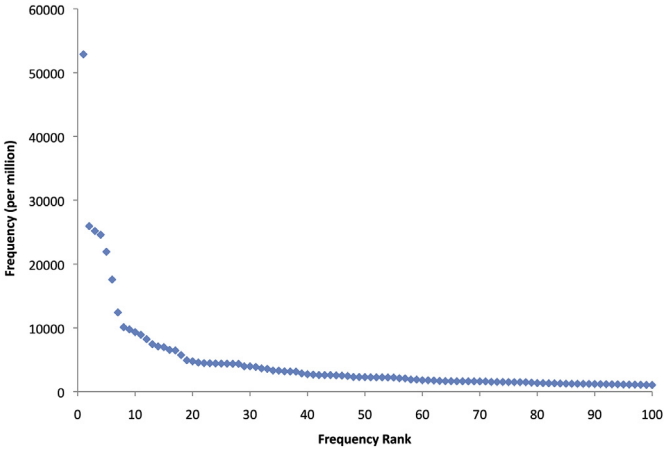

Fig. 1. The frequencies of the 100 most common words in the Corpus of Contemporary American English (Davies, 2009) plotted by rank.

Other language scientists have noticed at least some aspects of this problem and have argued that words do not "convey" or "transmit" meanings (Wittgenstein, 1952; Ramscar and Hahn, 1998; Harris, 1998; Love, 2004) or that phonemes cannot be the basic units of word specification (Faber, 1992; Linell, 2005; Port and Leary, 2005; Port, 2010). What has been lacking, however, are concrete proposals for how language might work in the absence of discrete building blocks, and without the idea that words and sentences etc. somehow transport meanings from one person to another.

Shannon's deductive model of communication. In contrast to the inductive models that have dominated linguistics—where learners and theorists seek to identify the units out of which larger form and meaning structures are built—the best formal models of both communication (information theory) and of how people learn (learning theory) describe deductive processes based on prediction and discrimination (Shannon, 1948; Kullback and Leibler, 1951; Rescorla and Wagner, 1972; Rescorla, 1988; Ramscar, 2010, 2013; Ramscar et al., 2010a; Ramscar and Baayen, 2013). Neither information theory nor discriminative learning theory adopts a constructive, compositional approach. Instead, both recast the problem as being that of *reducing uncertainty about the current state of the system* from among all possible states. Both learning and information theory adopt a discriminative (or deductive) approach. Neither seeks to "build" an understanding of a message, but rather each poses their respective problem as being that of eliminating alternatives. This is because both are scientific frameworks that seek to specify exactly *what* is learned or communicated. They do this without recourse to traditional psychological terms like "concept," "encode" and "meaning" whose use in vague and underspecified ways is widespread in descriptions of language and communication. Accordingly, although Shannon's (1948) statement of the problem of information theory for engineering is often misconstrued, his theory of information describes a communication problem that is different in critical ways from that embodied in traditional linguistic and psychological theory.

Shannon (1948, p. 623) states the central problem of communication as being "that of reproducing at one point either exactly or approximately a message selected at another point" and observes: "Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages." Many theorists influenced by traditional linguistics have mistakenly inferred from this that information theory *ignores* or leaves out the meanings that have traditionally been assumed to be "encoded" in linguistic signals. This is a misunderstanding. The semantic aspects of communication are irrelevant only to the engineering problems that his theory of artificial communication sought to solve. Semantics is not irrelevant to linguistic communication. That does not mean that information theory has nothing to teach us about language. Shannon was a key contributor to the development of coding theory, and thus acutely aware that the encoding of a message must contain a set of discriminable states that is greater than or equal to the number of discriminable states in the to-be-decoded message. It seems clear that Shannon also knew that English orthography lacks the coding resources required to discriminate the multifarious dimensions of meaning that traditional linguistic theories have supposed language must "encode" (see also Ramscar et al., 2010a).

Or, to put it another way, once the notion of encoding is formalized (e.g., Shannon, 1948), it is clear that natural languages lack the coding resources to encode all of semantics (Ramscar and Port, 2015; Wittgenstein, 1952). The semantic aspects of communication were irrelevant to Shannon's engineering purposes. Yet there are other important differences between natural and artificial communication. Many of the problems posed by natural communication must also be solved in order to facilitate artificial communication. Because we believe that information theory has many insights to offer when it comes to this problem and is often misunderstood (Shannon, 1956), we shall first describe Shannon's solutions for artificial communication purposes – focusing in particular on the deductive conceptual analysis Shannon used to frame these solutions – before describing how his analysis can be expanded to model natural communication.

5. Artificial communication – Shannon's system solution

The system solution (MacKay, 2003) to the communication problem that Shannon (1948) proposed comprises an information source and a destination and adds a transmitter before the communication channel and a receiver after it (Fig. 2). This is much as one might expect to find in any communication system (including traditional linguistic models that assume that meanings are encoded in linguistic signals). At the time Shannon developed information theory, a communication channel

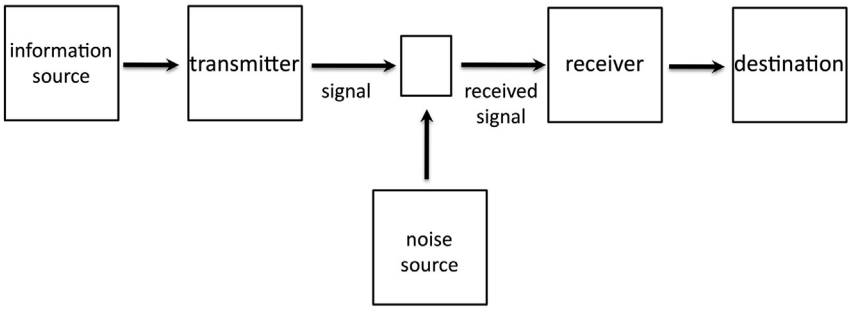

Fig. 2. A schematic diagram of an artificial communication system (based on Shannon, 1948).

usually comprised a simple wire over which continuous changes in a voltage signal were transmitted inevitably adding noise to the transmitted message. Thus a received message would comprise a mixture of the source message and noise. In order to enable the signal to be discriminated from the noise, the transmitter adds redundancy to the source message before transmission. This encoding process adds bits to the message (derived from the source message itself) that enable legitimate or possible messages to be discriminated from illegitimate or impossible ones. The decoder then uses the redundancy introduced by the encoding system to discriminate the original source signal from both the added noise and other possible signals. Note here that what encoding means in this context is simply the process of adding information that allows valid signals to be discriminated from invalid signals.

To facilitate this, the scope of possible and impossible messages is predefined in common *source* and *channel* codes, which can be combined (e.g., Soleymani and Khandani, 1991). In a Shannon system, it is essential that the transmitter and the receiver have access to the same codes. The *channel* code is designed to deal with the problem of channel noise. It defines the redundant information that enables a source message to be recovered even after its signal has been contaminated by noise in transmission. The *source* code defines the space of possible messages that can be transmitted in the system and is usually structured so as to allow the information in messages to be transmitted most efficiently.

Channel coding aims to maximize the discriminability of signals while minimizing the amount of redundancy that needs to be added (Hamming, 1950). Because signals are continuous and because the redundancy in a signal comes in the same form as a message, the distances between strings must be quantified in order for error detecting and error correcting codes to be specified. The discriminability of two strings of equal length can be quantified simply by counting the number of positions at which the corresponding symbols are different between the strings (called its Hamming distance). Hamming distances can be used to quantify the minimum number of errors that could have transformed one string into another. The Hamming distance between the bit strings 1000 and 0101 is 3 (because 3 bits must be flipped to transform 1000 into 0101), whereas the distance between 1000 and 1001 is 1. An analog that may be familiar to some linguists are Levenshtein (1966) distances, which characterize the difference in two words' orthographies as the minimum number of single-character edits (i.e., insertions, deletions or substitutions) required to change one into the other.

Coding and the distribution of information. Shannon's (1948) coding theorem proves that the discriminability and efficiency of a coding scheme can be maximized for efficiency if the distribution of the "words" used in the source language is highly skewed. The important idea behind Shannon's rather complex mathematical proof can be grasped by considering possible systems of personal names in a hypothetical community that are made up of strings of naming tokens that are encountered sequentially (Ramscar et al., 2013e, 2014). If all men are given a single name and if 33% of them are given the name type *John* and only 1% the name *Cornelius*, then *Cornelius* will provide a listener with more information than *John*. At the same time, *John* will be easier to remember than *Cornelius* (simply because if you guess someone is called *John* in this system, you will be right a third of the time). On the other hand, the name *Cornelius* will benefit from the fact that the *John*s don't have unique names. This distribution will result in a situation where picking a name at random from the list of all men's names will yield *Cornelius* far more often than its frequency predicts. This fact is important not only for the memorability of *Cornelius*, but will also influence the likelihood that the name will actually be "heard" in actual speech, since perception of the name is influenced by the degree to which the name is expected (Ernestus et al., 2002; Kemps et al., 2004).

Now, although a community with a single-name system of personal names would force parents to choose between the benefits of high and low probability names, typical name grammars code personal names in *strings*, that is, as sequences of name tokens. And if at each point in a string sequence, the set of tokens that can occur is distributed in the same way as we just saw above with English first names, communication across these sequences can be optimized for efficiency. Imagine a system in which name tokens occur in sequences, as first names and family names or surnames, then, if the *Johns* are given low probability surnames like *Farquar,* and if *Cornelius* is given a more likely surname such as *Smith*, then for the purposes of efficient discrimination of individuals, everyone is likely to benefit: A system so-arranged will allow each individual to be given a name that distinguishes them from their peers, while at the same time distributing information across the names so as to maximize the memorability of names and the efficiency of their processing in speech. By contrast, if everyone were given a single unique name (Ramscar et al., 2013e, 2014), the memorability and efficiency of processing names would be made far more difficult, because the unpredictability associated with each particular name would be maximized.

We can extend this example to illustrate how a skewed distribution can be combined with the manipulation of degrees of similarity between code words in order to further optimize signaling. In English, *John*, *William,* and *Thomas* have historically comprised around 50% of all male name tokens. Similarly, other lexical types have similar distributions, such as numbers, where *one, two,* and *three* account for around 50% of all number tokens. The names and the number words are Zipf distributed (where the frequency of a word is inversely proportional to its frequency rank). Similarly, in Korean, 50% of family names are *Kim, Lee* or *Park* (Ramscar et al., 2013e, 2014). Because the Levenshtein distances (measuring phonetic similarity) between *John*, *William,* and *Thomas* are large, as are those between the number words *one, two*, and *three* and between *Kim, Lee* and *Park* the possibility of confusing *John* with *William* or *Thomas* as well as *one, two* and *three* or *Kim, Lee* and *Park* is minimized. By contrast, if the most common male names were *John*, *Lon* and *Don* (where the Levenshtein distances are much smaller) a listener's ability to both detect and correct for errors when male names are encountered would be severely reduced.

Do natural languages employ a system solution? Thus it turns out that English names (and numbers, etc) have a distribution that resembles an optimal signaling code, as indeed, do the name systems of all the world's major languages (Ramscar et al., 2013e, 2014). It is notable that the Zipfian distributions (as shown in Fig. 1) that are found at every sequential choice point in language closely resemble the distributions that Shannon's (1948) source code theorem proves are optimal for

signaling purposes. Before we expand on this point, it is important to re-emphasize what Shannon's system solution to the problem of electronic signaling does and does not involve:

- The system solution requires that both the source encoder and the destination decoder share codes that pre-specify the scope of the possible messages that can be transmitted across the channel.
- The *meaning* of messages is irrelevant here. The goal of Shannon's system solution is that the receiver be able to successfully *construct* the source message (which is essentially a bit string) from the received message by *discriminating* the source message from other possible messages that might have been selected, and from the noise that is introduced by the communication channel;
- The decoder does not interpret or expand on the source message in any way. It simply allows the source message to be reproduced at the destination with no loss of signal content.

The communication process envisaged in information theory is deductive. The problem of communication over a noisy channel is solved by pre-specifying the space of possible messages in source and channel codes, and then employing these codes so that signals can be discriminated from noise by a process of elimination.

The difference between the inductive process envisaged by linguists and Shannon's system solution is hopefully clear: The latter is designed to allow the receiver to deduce the message encoded at the source with a high degree of accuracy. The use of the word "decode" in information theory does not correspond to the notion of a listener decoding a speaker's meaning in inductive models of language. This latter notion ignores the fact that words do not encode meanings in any way that corresponds to our normal understanding of "encode" (Ramscar et al., 2010a; Ramscar and Port, 2015).

Although meaning is irrelevant in a Shannon engineered system, it is clearly relevant in human communication. The idea that words encode meanings has strong intuitive appeal. But formal analysis of coding has determined that for every discriminable state that a receiver might possibly want to recover from a code, a corresponding discrimination must be present in the code. However, since it is clear words in conventional descriptions have far fewer discriminable states in their forms than there are discriminable states in their meanings, it also seems clear that technically, words cannot encode meaning. However, decoding simply corresponds to the receiver being able to successfully discriminate a communicated message from noise and from other possible messages. This brings us to the differences between human communication and electronic signaling: First, meaning *is* an important part of the shared code that makes human communication possible. And second, unlike the senders and receivers of artificial signaling systems, human senders and receivers are capable of learning, which is itself a discriminative process.

Accordingly, we suppose that a language, i.e., the source code of a human communication system, is a distribution of acoustic and semantic contrasts that are structured so as to facilitate the incremental reduction of semantic and acoustic uncertainty during speech. Since this system of contrasts considered as a code serves to reduce semantic uncertainty, semantics provides critical information to help identify acoustic contrasts (Ramscar et al., 2010a) and acoustic contrasts that have already been realized provide information about later contrasts. The message consists of both acoustic phonetic information and semantic cues relevant to the speaker's intention. Although engineered signaling systems are designed to maximize certainty in signaling, human communication can tolerate large degrees of uncertainty. This is both because learning is driven by uncertainty and because from a semantic point of view, almost all communication can be seen as serving the purpose of inducing learning in listeners (and speakers).

Consider the following dialog with certain time points marked with superscripts:

A *Whaddya wana* [1] *do to*[2]*ni*[3]*ght?* (What do you want to do tonight?)
B *I dunno. Whadda you wanna do?* (I don't know. What do you want to do?)
A [4]*I dunno. Wanna go to the*[5] *m*[6]*ovies*[?] (I don't know. Do you want to go the movies?)
B Sure. What [7] time?
A Se'[8]un? (Seven)

Traditional accounts of speech would first try to explain how the relevant acoustic phonetic units are identified, and then how these units are used to 'retrieve' meanings. By contrast, we suppose that A and B have learned a structured system of acoustic and semantic contrasts, such that by time-point [1], B has reduced her uncertainty about A's intentions sufficiently to understand that the message involves a question about her desires. Accordingly, she will be anticipating the contrast following [1] with a high degree of confidence because, based on both semantic and token probabilities, the set of contrasts (*do, see, eat…*) that are likely to occur here is small. Similarly, based on the expectations that B will have built up at time [2], it is likely at [3] that B will have eliminated any remaining semantic uncertainty (see Balling and Baayen, 2012) and successfully reconstructed A's message.

Thus at [4] B can anticipate the scope of A's proposal will be related to the set of things plausibly done in an evening. At [5], B can eliminate options conventionally preceded by an indefinite article (i.e., *a restaurant*) and thus, by [6], based on the context, B is likely to have fully reconstructed A's message, because all other likely contrasts will have been eliminated.

At [7], the articulation of *what* will increase the likelihood of the contrast *time* occurring (because although "*what do you want to see*?" might be anticipated here, the details of the articulation of *what* in that context is likely to differ – a glottal stop vs. a flap, see e.g., Jurafsky et al., 2001). And finally, given the small set of plausible contrasts possible following the semantic

cues available at [8], if B is an experienced English listener, she will be likely to 'hear' A articulate *seven* even if, in fact, this articulation is underspecified (Ernestus et al., 2002; Kemps et al., 2004).

Meanwhile, if B is a less experienced listener, she might be forced to request a clarification – *what?* – in the case of this underspecified articulation. Assuming that A supplies this clarification (rearticulating *seven* with more acoustic detail), B will learn from this experience and be better able to anticipate it in future (Ramscar et al., 2011).

Thus, crucially, the model of speech perception we describe does not rely on speakers and listeners having access to an inventory of discrete acoustic or semantic elements. Nor does it require that speakers' and listeners' representations of contrasts be identical (which, given the facts of vocabulary acquisition, is highly unlikely; see Ramscar et al., 2014; Keuleers et al., 2015). We noted above how experienced listeners can discriminate a plural from a singular form based on acoustic differences in the vowel (Kemps et al., 2005). It is clear that this discrimination has to be learned, and it is unlikely that younger speakers are sensitive to it. However, the presence of further discriminating information (the plural suffix itself) will enable a child to discriminate and reconstruct plural and singular nouns in an adult's message, facilitating successful communication despite the fact that a child and an adult will have somewhat different systems of cues and contrasts.

Instead of presuming that speakers and listeners share an inventory of units, the model we describe relies on their having access to a particular set of learning and processing mechanisms, and on a distribution of semantic cues and acoustic contrasts that form a system in which the confusability of contrasts is minimized in context. This way different messages can be discriminated in real time through the elimination of alternatives. Accordingly, we now review the evidence that supports this suggestion.

6. Psychological mechanisms that underpin language

The model we propose for language production and perception supposes that psychological mechanisms that enable the production and perception of language are fundamentally similar to those that underpin other human and animal behaviors. However, as well as borrowing from Shannon's insights, the approach is based on what may be one of the most important – and most underappreciated – discoveries of the past century: Learning does not lead to the acquisition of a simple inventory of isolated cues and their outcomes. Rather human and animal learning has been shown to be an inherently systematic process (Rescorla and Wagner, 1972) that discriminates the information in the environment that best allows an agent to predict the events that occur in it (Ramscar, 2010; Ramscar et al., 2010a,b, 2013a).

The result of the learning process is a system of knowledge in which the process of "understanding" can be seen as progressive reduction in an agent's uncertainty about environmental events, such as speech, that unfold in it. The high-dimensional space afforded by a mammalian brain can be represented as an input vector of all the sensory information available to an agent. This vector records evidence that enables the agent to discriminate which state (of the states it has learned to discriminate) is most likely to occur next. Learning shapes that vector by emphasizing (via larger weights) or deemphasizing (with smaller or negative ones) aspects that aid or hinder the prediction of future events. Understanding is thus an active process that unfolds in time relative to a learned system of contrasts. Where a learned contrast is anticipated well enough that the level of uncertainty experienced is low, understanding is good. Where uncertainty is high, a learner finds the situation more difficult to understand.

Applied to speech, the process of discrimination of acoustic and semantic cues will be aided if the perceptual cues to semantically similar events occupy regions in the speech vector space that are far apart. For example, a listener's ability to detect the errors that drive the discrimination of the most frequent number words will be increased if the words are distinct in their acoustic phonetics. Thus, it helps that the word for 1 item is *one,* for 2 items *two,* and for 3 items *three.* And it would be impaired if 1 were *one,* 2 *sun* and 3 *fun* (see Ramscar et al., 2011). Any errors that are detected will be reflected in the learner as changes in the system of weights connecting the network cues (vector points) and outcomes in the system (see Ramscar et al., 2010a; Ramscar et al., 2013a for reviews). Whenever a cue is followed by a predicted outcome, the weight on a link will be increased, and whenever the predicted outcome does not occur, the weight will be reduced.

To illustrate how this inherently systematic process can produce unexpected results in relation to language, we will consider how discrimination learning influences the relations between items in a maximally simple lexicon by considering how Paired-Associate Learning (PAL) – a research paradigm with a long history in psychology –reflects changes across the lifespan. In PAL tasks, participants are presented a sequence of word pairs, *w1* and *w2.* After presentation of a list of pairs, participants are prompted with *w1* and asked to respond with *w2*. It has been shown that older participants do worse at this task than younger ones (des Rosiers & Ivison, 1986). Indeed, this is the case even when 30–39-year-olds are compared with 20–29-year-olds (Fig. 3).

At first glance, it seems that those in their twenties perform quite a bit better. However, if the pairs of items are sorted by their overall difficulty (as they are in Fig. 3), an interesting pattern emerges: 30–39-year-olds do only slightly worse than the younger group on what we might call meaningful pairs, such as *baby–cry* and *North-South*. However, the difference in performance between the two groups increases as the meanings of the word pairs becomes more unrelated, e.g., *crush–dark* and *jury-eagle*, such that the pairing become more meaningless.

These data illustrate the systematic nature of learning systems (Ramscar et al., 2013a). If we were to focus on the pooled data across all pairs, we might conclude that because they learn more items, the 20–29-year-olds do better at this quasi-linguistic task than the 30–39-year-olds. We might go so far as to conclude, as some have done, that learning in the 30–39-year-olds is somehow compromised (see e.g., Salthouse, 2011). However, if we attend to the entire system, we notice how

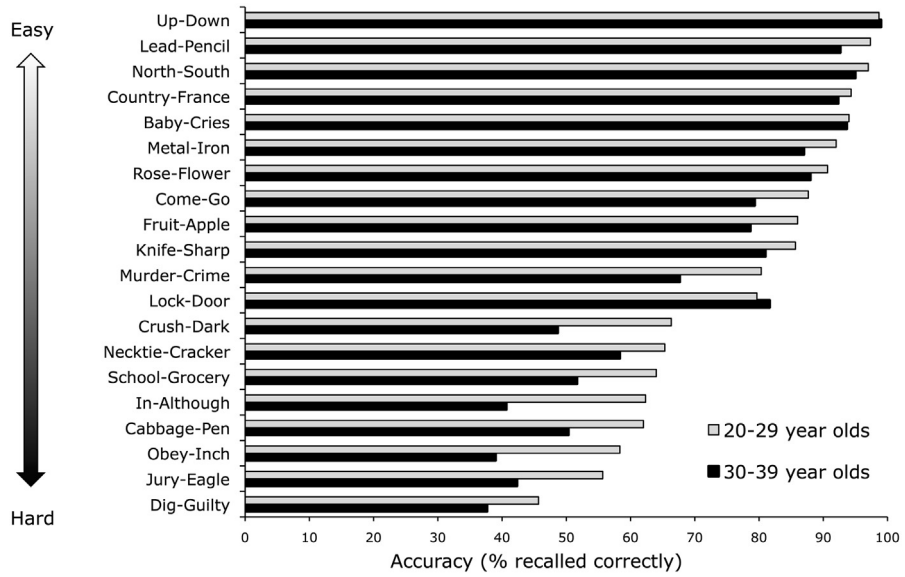

Fig. 3. Average by-item performance for 400 adults aged 20–29 and 30–39 (50% females per group) on forms 1 & 2 of the WMS-PAL subtest (data from des Rosiers & Ivison, 1986).

the discrepancy between meaningless and meaningful pairings is greater for the older participants, and might conclude that the lexical system is less well learned by the younger participants. Learning models can be used to determine which of these conclusions, that younger participants know less about the language or that older ones are somehow impaired, is more likely.

To see the reasoning, it is important to stress that although noticing which events co-occur is a useful heuristic for informativity, it is far from sufficient as learning a mechanism. Consider two cues that regularly co-occur as events, say, the paired associates: *the–girl* and *little–girl*. If the word *the* occurs more often when not followed by *girl*, than *little* occurs without *girl*, then *little* will be more informative as a predictor of *girl* than is *the*. The Rescorla-Wagner learning mechanism will reinforce the value of *little* and *the* for *girl* each time they co-occur, but will decrement them each time *little* or *the* is not followed by *girl*. This mechanism in effect evaluates cues according to the degree to which they reduce uncertainty – learning more when uncertainty is high, and less when uncertainty is low. This means that when events are well predicted, little learning takes place (called blocking). Accordingly, because the rates at which *the* and *little* occur without *girl* differ, over time the value of *little* for *girl* will come to outweigh that of *the* for *girl.*

To illustrate how isolated instances of learning scale up within a system, we can now return to the word pairing data in Fig. 3. The development of large on-line text corpora allow linguistic parameters to be estimated for this kind of learning, and permit testing using formal learning models (Danks, 2003) to account for the data. We can apply these parameters to the data above by estimating (a) co-occurrence from the frequency at which the two words occur near each other, (b) background rate from the frequency of the first word, and (c) blocking from the ratio of the frequency of the second word as compared to the first (Ramscar et al., 2013c). In a simple linear regression to predict the relative performance of the two groups, these standard predictors of associative learning account for over 85% of the variance between the easy and hard pairings for both the 20–29-year-olds and the 30–39-year-olds. This suggests that the differences in their responses to these items do not reflect differences in learning ability. Indeed, the model shows they reflect exactly the same learning system, but that experience shapes the performance of that system and experience differs as a function of age (Ramscar et al., 2013c).

These results show that although it might seem intuitively that learning to pair *jury* with *eagle* is a combinatoric process involving only the words *jury* and *eagle*, in fact, *jury* and *eagle* are part of a system, and the relationship between *jury* and *eagle* is determined by the properties of the entire system. When more experienced speakers of English are asked to learn to pair *jury* and *eagle,* they have to do so using a memory system in which *jury* and *eagle* inhibit each other. That is, the contrast between words that are associated and words that are not associated is clearer. Accordingly, this prior learning has to be reversed in order for them to recall *eagle* given *jury*.

Linguistic communication as uncertainty reduction. Historically, theoretical approaches to human communication have overwhelmingly adopted a generative, taxonomic approach to explaining how meaning is communicated in text or speech. Linguistic communication has been seen as a process of encoding, transmitting and decoding tokens of meaning types. These meaning types have been assumed to be taxonomically organized, and encoded and decoded by rules that allow messages to be generated from them. From this perspective, the challenge facing both language learners and theoretical linguists is inductive. The correct taxonomy of meaning types and generative rules for a given language must be inferred from whatever data is available to the learner or theorist. However, both history and logic suggest that the problems these inductive challenges pose may be insoluble (Ramscar and Port, 2015).

As our PAL example illustrates, both information theory (the study of artificial communication systems) and empirically grounded psychological theories of learning describe deductive processes based on prediction and discrimination (Shannon, 1948; Kullback and Leibler, 1951; Rescorla and Wagner, 1972; Ramscar et al., 2010a, 2011; Baayen, 2011; Baayen et al., 2011). These results suggest communication need not require the induction of taxonomic semantic classes and rules for their combination. Both learning and information theory adopt a discriminative approach. Neither seeks to "build" an understanding or a message. Rather each poses their respective problem as being that of eliminating alternatives. Learning and information theory thus follow Sherlock Holmes' dictum: "When you have eliminated the impossible, whatever remains, however improbable, must be the truth."

The semantics of a discriminative language system is very different from other systems of semantics. It works because, as with other cognitive processes, linguistic processing proceeds in a predictive, discriminative fashion. As a language evolves over many generations, it develops a system of semantic and acoustic contrasts that provide speakers with structured expectations about the form that linguistic signals can take. Linguists describe these form patterns as "words, morphemes, phonemes, phrases, idioms, constructions, intonation patterns" etc. Learning these relations means the learner must discriminate the relevant conventional phonetics (that is, the relevant acoustic events) from all other phonetic patterns, and discriminate the semantics associated with the various phonetic patterns afforded by the system (Ramscar and Port, 2015). The discriminative learning process assures (as illustrated by the PAL example) that each cue has systematic relationships with many others. In a communicative exchange, the speaker uses their own model of the linguistic code to assemble a signal sequence. The listener is able to reconstruct the speaker's message because at each point in time, this signal will provide the listener with information that systematically reduces the listener's uncertainty about the nature of the intended message. If all goes well, the intended message and the received message will align – up to the limits of the semantic discriminations of the signals employed.

In an artificial communication system, unlike natural language, the source code specifies the absolute scope of what those possible alternatives can be, because the goal of a communication system is to use the code to restore a message from the noisy signal with near certainty as it unfolds in real time. In the light of the overwhelming evidence for prediction at every level of language processing (for reviews, Ramscar et al., 2010a,b), this in turn highlights an important difference between natural and artificial communication systems as well as the clear connection between communication and learning: Consider a listener who could predict with absolute certainty each and every aspect of the signal produced by a speaker as it unfolds in real time. We might say that, in this case, the speaker's message, by being perfectly predictable, was completely uninformative. However, it is rare that speakers are perfectly predictable (unless they are, for example, chanting a prayer). Rather, because speakers and listeners embody learning systems, natural communication is "meaningful" to the degree that listeners are not able to predict speakers perfectly. The goal of almost all speech acts is to effect a reduction in a listeners' uncertainty and to see that a listener learns something from them.

This is to say that comprehension is not a combinatorial process of building up a representation of an understanding. Rather it is a time and sequence dependent process that reduces a listener's uncertainty about what a speaker means. An attractive feature of this approach is that it allows exactly the same analysis to be applied to *Hello!* (where the tone and demeanor of a speaker reduce the listener's uncertainty about the way the rest of a conversation may go) as it does to *The boy kicked the ball* and even *Her father threw a ball* (where the distributional properties of English mean that the meaning of the latter is far more uncertain than the former given no other context). It is also why Dutch speakers hear the sound *[eik]* as meaningless without context but in context hear *eigenlijk,* "in fact".

As the foregoing indicates, treating learning as a systematic process and language processing as a function of learning highlights the importance of the actual distributional properties of language. A language is not learned as a simple inventory of form-meaning "associations." It is learned discriminatively in context (Ramscar et al., 2013a) and serves as a complex system that regulates the way contrasts are experienced in context. Thus, the fact that language is Zipf-distributed at every conceivable level of description across the world's languages becomes a highly relevant empirical fact, as does the specific distribution of lexical contrasts within any given level of description.

What is notable from a speech perspective is the degree to which the empirical distribution of English maximizes the contrast of the majority of tokens that can be expected to occur whenever a name or a number can be expected to occur. This serves to minimize uncertainty in many contexts. It seems likely that the distribution of the system of forms and meanings in any language has been shaped by communicative selection pressures acting on generations of speakers and listeners. These distributions facilitate both language learning and efficient communication.

Learning lengthy patterns. Further support for these ideas comes from two main sources: artificial language learning experiments (Ramscar, 2002, 2010, 2013; Ramscar et al., 2010a, 2013b,c) and studies of large text corpora (Ramscar et al., 2010a, 2013d,e, 2014; Baayen et al., 2013). An example of the latter that is pertinent here is provided by Baayen et al. (2013) who show how a learning model can simulate the results of an experiment demonstrating listeners' sensitivity to frequency effects on 4-word phrases (Arnon and Snider, 2010; see also Bannard and Matthews, 2008) even though the model itself contains no explicit representation of 4-word phrases.

It is well-known that language users are sensitive to the frequency of many supposed linguistic units: word frequency, sound-combination frequency and many other linguistic components (for review, see Monsell, 1991). The effect is that listeners can make decisions about text fragments much more quickly for frequent words and phrases than for less frequent ones. To examine the degree to which people store large "exemplar chunks" of language, Arnon and Snider (2010) measured participant reaction-times to allowable English phrases (such as *Out of the house*) versus those that were not allowable (like, e.g., *Out the of*

house). The study employed a set of 47 target words embedded in 4-word text phrases that were balanced for the corpus frequency of component words but which differed in the frequency of occurrence of the 4-word sequences (that is, the 4-grams) in a large corpus of text. The critical feature is that each possible phrase was paired with another partner where one 4-gram of the pair was 3–10 times as frequent as the other (such as, in this case, *Out of the game* as the less frequent partner of *Out of the house*).

Surprisingly (given many theoretical accounts of language processing) the more frequent phrases were responded to faster than the perfectly natural but less frequent phrases. This is a surprise because at first blush, one might think that in order to respond faster to the more frequent phrases, speakers would have to somehow store these phrases as exemplars and somehow count their occurrences.

An alternative explanation of this phenomena is offered by Baayen et al., (2013). They simulated this speech task with an analogous reading task using their program called the Naïve Discriminative Reader (Baayen et al., 2011), a learning model that seeks to simulate the way people predict lexical contrasts from orthographic cues in reading. To examine whether their model could simulate these results, they took the set of test phrases used by Arnon and Snider and then located all of the 4-word sequences containing Arnon and Snider's target words in the *British National Corpus* database (563,000 tokens of 337,000 types). Baayen et al. then trained their model to use only single letters and letter pairs as cues to predict the lexical contrasts in each of the 4-word phrase tokens. The n-grams were trained to predict each phrase in its entirety, i.e., the orthographic cues to the lexical contrasts *out, of, the,* and *game* were all presented simultaneously, and the model itself then used discrimination learning to establish which cues best predicted which contrast. After training, the activations of the set of lexical contrasts in each phrase resulting from the presentation of the relevant orthographic cues to each phrase successfully predicted the faster reaction times for the higher-frequency and slower RTs for lower-frequency phrases in Arnon and Snider (2010).

This system exhibits many notable properties. The only "units" it employs are n-grams (in this case just bigrams and unigrams of letters) and its inputs produce a continuous range of activations across the contrasts upon which it trained. The system stores neither whole words, nor word-bigrams, let alone 4-word phrases, yet despite the apparently impoverished representations it employs, it is able to predict empirical responses to 4-word phrases. It is thus worth dwelling for a moment on how exactly it is that the model is able to do this. Despite whatever qualms one might have about English orthography, this result demonstrates empirically that when a biologically plausible learning model is used to simulate the acquisition of the relationships between orthography and lexical contrasts in context, a representative sample of the actual distribution of orthographic n-grams and lexical contrasts in English contains sufficient information to allow readers to discriminate the relevant lexical contrasts contained within that sample with high levels of confidence. If the distribution did not contain this information, the simulation would not work. Meanwhile, along with many other successful simulations of empirical phenomena by the same model (or variants of it, Baayen et al., 2011, 2013; Ramscar et al., 2013c, 2014; Baayen & Ramscar, 2015) the success of its empirical fit lends strong support to the idea that people actually do learn to discriminate the orthographic cues to the contrasts they make in reading in this way. And, of course, it also supports the possibility that speakers may employ similar acoustic and visual cues for predicting stretches of up-coming speech.

Language change and language learning. Languages evolve as they are used. Members of a speech community gradually shape the language so as to (a) maximize the ability of listeners to discriminate the linguistic signals (Lindblom, 1990), (b) reduce the articulatory effort required of speakers for communication, and also (c) shape the inventory of linguistic contrasts so as to reduce listener uncertainty across the ever expanding range of semantic contrasts that are relevant to that community. Across a language, and across the lifespan of speakers (because there is no endpoint for language learning in the lifespan, see Ramscar et al., 2014; Keuleers et al., 2015), the results of this shaping create systems of contrast at multiple levels of granularity that are often characterized descriptively as the units called distinctive features, phonemes, morphemes, words, etc. We claim however that they are not units to be manipulated by themselves but simply regularities that are the result of shaping the signals of language to make them more efficient.

This process also produces distributions of contrasts that are highly skewed (Zipf, 1949; Baayen, 2001), and within these distributions, there is a tendency for the most frequent contrasts at each level of description to be suppletive (Blevins et al., 2015). Thus frequent forms (e.g., *mouse*/*mice, see*/*saw* and *eat*/*ate*) tend to be irregular because their form-meaning relations are more fully discriminated: The plural form *mice* is more discriminated from other plural forms as compared to, say, *rats* and *pots* (see Ramscar and Yarlett, 2007; Ramscar et al., 2013a).

One constraint that appears to influence the distribution of regular and irregular forms is the trajectory of human brain development, and the way that this development influences the nature of learning across childhood. Compared to other primates, human cortical development is uneven. Synaptogenesis in the visual and auditory cortex peaks just a few months after birth, while the same development occurs much later in the prefrontal cortex, which doesn't fully mature until late adolescence (Ramscar and Gitcho, 2007; Thompson-Schill et al., 2009). An important behavioral consequence of delayed prefrontal development is that young children have less ability to selectively attend, or to engage in behaviors that conflict with prepotent responses (Ramscar et al., 2013d). In adults, prefrontal control mechanisms bias responses and attention according to specific goals or context, selectively maintaining task-relevant information and discarding task-irrelevant information (Ramscar and Gitcho, 2007; Thompson-Schill et al., 2009). The absence of this capacity in young children can be illustrated by contrasting their performance with that of adults on biased selection tasks, such as guessing the hand an M&M is in, where the hands are biased 25:75. Children up to age 5 tend to overmatch, fixating on the higher-probability "good" hand. After age 5, however, a probability matching strategy emerges. This is a rare instance in which children's inability to

think flexibly is an advantage (since probability matching actually reduces the number of M&Ms won by children over 5 years old and adults; Thompson-Schill et al., 2009).

The cognitive flexibility that comes with cognitive maturation appears to be disadvantageous whenever learning is best pursued in a purely bottom up fashion (Ramscar et al., 2013a). Such a strategy favors the learning of arbitrary conventions, such as the irregular aspects of even simple morphological systems like those found in English (Ramscar et al., 2013a). Linguistic knowledge is, in its essence, conventional: In the presence of a linguistic cue, a social animal needs to be able to understand or respond appropriately given the context. For this to happen, linguistic signals, must be both conventional and internalized (Wittgenstein, 1952). Learning the system cues that yield appropriate understanding for all contrasts (but especially those which are unpredictable or irregular) is far more likely to happen if learners are unable to filter their attention too much during the course of learning. Given a similar set of cues and labels to learn, any two young learners each tend to sample and learn from the environment in much the same way as one another. This in turn makes it more likely that they will come to have similar expectations about the contrasts that they acquire, and the contexts in which they learn them (Ramscar et al., 2013a,b; see also Hudson Kam & Newport, 2005, 2009; Friederici et al., 2013; Finn et al., 2013, 2014).

These changes offer a biological explanation for the widely accepted idea that intention reading or social prediction is an important component of language learning (Tomasello, 2008). The approach we have described simply extends intention reading to language processing more generally. Comprehension arises out of what a listener knows that makes a speaker predictable and from what a listener is able to learn from the speaker. Identifying the contrasts and constructions that a speaker actually employs leads to learning about why a speaker made the choices they did. In this regard, it is notable that the actual functionality of young children's learning capacities, and the delayed development of a more agentive and adult-like learning style appears to be well adapted for the acquisition of conventionalized predictive codes like language.

7. Conclusion

This paper has been concerned with two primary issues regarding language comprehension: units and processing. Units – distinctive features, phones, phonemes, words and morphemes – have traditionally been seen as central components of human language. Our observation is that speech in any language has many characteristics that roughly resemble these units. These regularities reflect the usage of a language over many generations to make it more effective for communication. These regularities have been highlighted and formalized in our practical orthographic systems such that we have strong intuitions favoring them as "the units of language". Because of their salience linguists and others have focused attention on them, and mistakenly have taken these units to be just the units that listeners use to construct syntactic parses and semantic structures.

We have argued, however, that although these units may be useful as tools in describing speech, they play at most a very small role in the real-time perceptual processing of language. Rather, we suggest that a young language learner, faced with someone uttering syllables at them, uses the same kind of learning system our mammalian cousins use – a discriminative learning system that attempts to predict what will happen next by learning what cues enable correct predictions about lexical contrasts and their semantic correlates as they unfold in continuous time. Because these cues and contrasts are discovered and refined as the result of experience, the granularity of both cues and contrasts are found in a wide range of sizes over a wide range of timescales, and will differ from person to person as well as from language to language.

Thus the first goal of this paper was to review the numerous reasons (no doubt largely familiar to many readers) why notional linguistic units like phonemes, words and morphemes are problematic when it comes to building a compositional account of spoken language. The second was to sketch an alternative account of language processing that avoids the apparently insoluble problems involved in identifying the compositional constituents of language. We would be the first to acknowledge that many details of this account require further development. However, our belief is that this alternative is both plausible and viable. We hope that at least some of our colleagues may be inspired to explore their own questions along these lines.

References

Arnon, I., Snider, N., 2010. More than words: frequency effects for multiword phrases. J. Mem. Lang. 62, 67–82.
Baayen, R.H., 2001. Word Frequency Distributions. Kluwer, Amsterdam.
Baayen, R.H., 2011. Corpus linguistics and naive discriminative learning. Braz. J. Appl. Ling. 11, 295–328.
Baayen, H., Milin, P., Filipović Đurđević, D., Hendrix, P., Marelli, M., 2011. An amorphous model for morphological processing in visual comprehension based on nadve discriminative learning. Psychol. Rev. 118, 438–481.
Baayen, H., Hendrix, P., Ramscar, M., 2013. Sidestepping the combinatorial explosion: an explanation of n-gram frequency effects based on native discriminative learning. Lang. Speech 56, 329–347.
Baayen, R.H., McQueen, J.M., Dijkstra, T., Schreuder, R., 2003. Frequency effects in regular inflectional morphology: revisiting Dutch plurals. In: Baayen, R.H., Schreuder, R. (Eds.), Morphological Structure in Language Processing. Mouton de Gruyter, Berlin, pp. 355–370.
Baayen, H., Ramscar, M., 2015. Abstraction, storage and naive discriminative learning. In: Dawbroska, E., Divjak, D. (Eds.), Handbook of Cognitive Linguistics. De Gruyter Mouton.
Balling, L., Baayen, R.H., 2012. Probability and surprisal in auditory comprehension of morphologically complex words. Cognition 125, 80–106.
Bannard, C., Matthews, D., 2008. Stored word sequences in language learning the effect of familiarity on children's repetition of four-word combinations. Psychol. Sci. 19 (3), 241–248.
Blevins, J., P. Milin & M. Ramscar (ms, 2015) Zipfian discrimination.
Bolinger, D., 1975. Aspects of Language, second ed. Harcourt, Brace Jovanovich.
Browman, C., Goldstein, L., 1995. Dynamics and articulatory phonology. In: Port, R., van Gelder, T. (Eds.), Mind as Motion: Explorations in the Dynamics of Cognition. MIT Press, Cambridge, MA., pp. 175–193.

Chomsky, N., Miller, G., 1963. Introduction to the formal analysis of natural languages. In: Luce, R.D., Bush, R.R., Galanter, E. (Eds.), Handbook of Mathematical Psychology, vol. 2. Wiley, New York, NY, pp. 323–418.
Chomsky, N., Halle, M., 1968. The Sound Pattern of English. Harper and Row.
Church, K.W., Gale, W.A., 1995. Poisson mixtures. Nat. Lang. Eng. 1 (02), 163–190.
Connine, C., Clifton, C., Cutler, A., 1987. Effects of lexical stress on phonetic categorization. Phonetica 44, 133–146.
Cutler, A., 2012. Native Listening: Language Experience and the Recognition of Spoken Words. MIT, Cambridge.
Danks, D., 2003. Equilibria of the Rescorla-Wagner model. J. Math. Psych. 47 (2), 109–121.
Davies, M., 2009. The 385+ Million word corpus of contemporary American English (1990-present). Int. J. Corpus Linguist. 14, 159–190.
des Rosiers, G., Ivison, D., 1986. Paired-associate learning: normative data for differences between high and low associate word pairs. J. Clin. Exp. Neuropsychol. 8, 637–642.
Erman, B., Warren, B., 2000. The idiom principle and the open choice principle. Text Talk 20, 29–62.
Ernestus, M., Baayen, H., Schreuder, R., 2002. The recognition of reduced word forms. Brain Lang. 81 (1), 162–173.
Faber, A., 1992. Phonemic segmentation as epiphenomenon: evidence from the history of alphabetic writing. In: Lima, S.D., Noonan, M., Downing, P. (Eds.), The Linguistics of Literacy: Typological Studies in Language, vol. 21. Johns Benjamins, Amsterdam, pp. 111–134.
Fant, G., 1959. The acoustics of speech. In: Cremer, L. (Ed.), Proceedings of the Third Intl Congress on Acoustics, Stuttgart. MIT Press, Cambridge (Elsevier, 1961). Reprinted in G. Fant (ed.) (1973) *Speech Sounds and Features*.
Finn, A.S., Lee, T., Kraus, A., Kam, C.L.H., 2014. When it hurts (and helps) to try: the role of effort in language learning. PloS One 9 (7), e101806.
Finn, A.S., Kam, C.L.H., Ettlinger, M., Vytlacil, J., D'Esposito, M., 2013. Learning language with the wrong neural scaffolding: the cost of neural commitment to sounds. Front. Syst. Neurosci. 7.
Friederici, A.D., Mueller, J.L., Sehm, B., Ragert, P., 2013. Language learning without control: the role of the PFC. J. Cogn. Neurosci. 25 (5), 814–821.
Gafos, A., Goldstein, L., Côte, M., Turk, A., 2012. Organization of phonological elements. In: Cohn, A., Huffman, M., Fougeron, C. (Eds.), Oxford Handbook of Laboratory Phonology. Oxford Univ. Press.
Gahl, S., 2012. Why so short? Competing explanations for variation. In: Gahl, S., Yao, Y., Johnson, K. (Eds.), Proceedings of the 29th West Coast Conference on Formal Linguistics. Cascadilla Proceedings Project, Somerville MA, pp. 1–10.
Hamming, R.W., 1950. Error detecting and error correcting codes. Bell Syst. Tech. J. 29 (2), 147–160.
Harris, R., 1998. An Introduction to Integrational Linguistics. Pergamon, Oxford.
Hay, J., Baayen, H., 2005. Shifting paradigms: gradient structure in morphology. Trends Cogn. Sci. 9, 342–348.
Honorof, D., Feldman, L., 2006. The Chinese character in psycholinguistic research: form, structure and the reader. In: Li, Ping, Hai Tan, Li, Bates, E., Tzeng, O. (Eds.), The Handbook of East Asian Psycholinguistics, Volume 1: Chinese. Cambridge U. P., pp. 195–208.
Hudson Kam, C.L., Newport, E.L., 2005. Regularizing unpredictable variation: the roles of adult and child learners in language formation and change. Lang. Learn. Develop. 1, 151–195.
Hudson Kam, C.L., Newport, E.L., 2009. Getting it right by getting it wrong: when learners change languages. Cogn. Psychol. 59 (1), 30–66.
International Phonetic Association, 1949. Principles of the International Phonetic Association. Cambridge U. P. (June 28, 1999).
Jakobson, R., Fant, G., Halle, M., 1952. Preliminaries to Speech Analysis: the Distinctive Features and Their Correlates. Technical Report 13. Acoustics Laboratory, MIT, Massachusetts.
Joos, Martin, 1948. Acoustic phonetics. Language 24 (2), 5–136.
Jurafsky, D., Bell, A., Gregory, M., Raymond, W., 2001. Probabilistic relations between words: evidence from reduction in lexical production. In: Bybee, J., Hopper, P. (Eds.), Frequency and the Emergence of Linguistic Structure. John Benjamins, Amsterdam, pp. 229–254.
Katz, S.M., 1996. Distribution of content words and phrases in text and language modelling. Nat. Lang. Eng. 2 (01), 15–59.
Kemps, R., Ernestus, M., Baayen, R.H., Schreuder, 2004. Processing reduced word forms: the suffix restoration effect. Brain Lang. 90 (1), 117–127.
Kemps, R., Ernestus, M., Schreuder, Baayen, R.H., 2005. Prosodic cues for morphological complexity: the case of Dutch plural nouns. Mem Cognit 33 (3), 430–446.
Keuleers, E., Stevens, M., Mandera, P., Brysbaert, M., 2015. Word knowledge in the crowd: measuring vocabulary size and word prevalence in a massive online experiment. Quart. J. Exp. Psychol., 1–28 (ahead-of-print).
Kim, I., Park, C., Lee, K., Kim, N., Lee, J., Kim, J., Lane, I., 2015. Development of highly accurate real-time large scale speech recognition system. In: 2015 IEEE International Conference on Consumer Electronics. IEEE, pp. 493–496.
Klatt, D., 1976. Linguistic uses of segmental duration in English: acoustic and perceptual evidence. J. Acous. Soc Amer 59, 1208–1221.
Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Annals Math. Stat. 22 (1), 79–86.
Lachs, L., McMichael, K., Pisoni, D.B., 2003. Speech perception and implicit memory: evidence for detailed episodic encoding of phonetic events. In: Bowers, J., Marsolek, C. (Eds.), Rethinking Implicit Memory. Oxford U.P, pp. 215–235.
Levenshtein, V.I., 1966. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Phys. Dokl. 10 (8), 707–710.
Lindblom, B., 1990. Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle, W., Marchal, A. (Eds.), Speech Production and Speech Modeling. Kluwer, Dordrecht, pp. 403–439.
Linell, P., 2005. The Written Language Bias in Linguistics: Its Nature, Origins and Transformations. Routledge.
Love, N., 2004. Cognition and the language myth. Lang. Sci. 26, 525–544.
MacKay, D.J.C., 2003. Information Theory, Inference and Learning Algorithms. Cambridge U. P., Cambridge, England.
Monsell, S., 1991. The nature and locus of word frequency effects in reading. In: Besner, D., Humphries, G. (Eds.), Basic Processes in Reading: Visual Word Recognition. Erlbaum, Hillsdale., NJ, pp. 148–197.
Öhman, S., 2002. Phonetics in a literal sense: 1. Do the letters of the alphabet have a pronunciation? Proc. Fon. TMH-QPSR Q. Prog. Rep. KTH 44 (1), 125–128.
Palmeri, T.J., Goldinger, S.D., Pisoni, D.B., 1993. Episodic encoding of voice attributes and recognition memory for spoken words. J. Exptl. Psych. Learn. Mem. Cogn. 19, 309–328.
Port, Robert, 1979. The influence of tempo on stop closure duration as a cue to voicing and place. J. Phon. 7, 45–56.
Port, Robert, 1981. The 'applied suffix' in Swahili. Stud. Afr. Linguist. 12, 71–82.
Port, Robert F., 2010. Language as a social institution: why phonemes and words do not live in the brain. Ecol. Psychol. 22, 304–326.
Port, Robert, Leary, Adam, 2005. Against formal phonology. Language 81, 927–964.
Port, R., Cummins, F., McAuley, D., 1995. Naive time, temporal patterns and human audition. In: Port, R., van Gelder, T. (Eds.), Mind as Motion: Explorations in the Dynamics of Cognition. MITP, Cambridge, MA, pp. 339–372.
Potter, R.K., Kopp, A.G., Green, H.C., 1947. Visible Speech. Van Nostrand, New York.
Ramscar, Michael, 2002. The role of meaning in inflection: why the past tense does not require a rule. Cogn. Psychol. 45, 45–94.
Ramscar, M., 2010. Computing machinery and understanding. Cogn. Sci. 34 (6), 966–971.
Ramscar, M., 2013. Suffixing, prefixing and the functional order of regularities in meaningful strings. Psihologija 46, 377–396.
Ramscar, Michael, Hahn, Ulrike, 1998. What family resemblances are not: the continuing relevance of Wittgenstein to the study of concepts and categories. In: Proceedings of the 20th Annual Conference of the Cognitive Science Society. University of Wisconsin – Madison.
Ramscar, Michael, Baayen, H., 2013. Production, comprehension and synthesis: a communicative perspective on language. Front. Psychol. 4, 233–236.
Ramscar, M., Dye, M., Popick, H.M., O'Donnell-McCarthy, F., 2011. The enigma of number: why children find the meanings of even small number words hard to learn and how we can help them do better. PLoS One 6 (7), e22501. http://dx.doi.org/10.1371/journal.pone.0022501.
Ramscar, M., Dye, M., McCauley, S.M., 2013a. Error and expectation in language learning: the curious absence of *mouses* in adult speech. Language 89, 760–793.
Ramscar, M., Dye, M., Klein, J., 2013b. Children value informativity over logic in word learning. Psychol. Sci. 24 (6), 1017–1023.

Ramscar, M., Hendrix, P., Love, B., Baayen, H., 2013c. Learning is not decline: the mental lexicon as a window into cognition across the lifespan. Ment. Lex. 8 (3), 450–481.
Ramscar, M., Dye, M., Gustafson, J.W., Klein, J., 2013d. Dual routes to cognitive flexibility: learning and response conflict resolution in the dimensional change card sort task. Child Dev. 84 (4), 1308–1323.
Ramscar, M., Dye, M., Hubner, M., 2013e. When the fly flied and when the fly flew: how semantics can make sense of inflection. Lang. Cogn. Process 28 (4), 468–497.
Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., Baayen, R.H., 2014. The myth of cognitive decline: non-linear dynamics of lifelong learning. Topics Cogn. Sci. 6, 5–42.
Ramscar, M., Yarlett, D., Dye, M., Denny, K., Thorpe, K., 2010a. The effects of Feature-Label-Order and their implications for symbolic learning. Cogn. Sci. 34 (6), 909–957.
Ramscar, M., Matlock, T., Dye, M., 2010b. Running down the clock: the role of expectation in our understanding of time and motion. Lang. Cogn. Process 25 (5), 589–615.
Ramscar, M., Gitcho, N., 2007. Developmental change and the nature of learning in childhood. Trends Cogn. Sci. 11 (7), 274–279.
Ramscar, M., Port, R., 2015. Categorization (without categories). In: Dawbroska, E., Divjak, D. (Eds.), Handbook of Cognitive Linguistics. De Gruyter Mouton.
Ramscar, M., Yarlett, D., 2007. Linguistic self-correction in the absence of feedback: a new approach to the logical problem of language acquisition. Cogn. Sci. 31, 927–960.
Rescorla, R., Wagner, A.R., 1972. A theory of Pavlovian condition: variations in the effectiveness of reinforcement and nonreinforcement. In: Black, A., Prokasy, W. (Eds.), Classical Conditioning II: Current Research and Theory. Appleton-Century Crofts, New York, pp. 64–99.
Rescorla, R., 1988. Pavlovian conditioning: it's not what you think it is. Am. Psychol. 43, 151–160.
Salthouse, T., 2011. Consequences of age-related cognitive declines. Ann. Rev. Psychol. 63, 5.1–5.26.
Saussure, F., 1916. In: Bally, C., Sechehaye, A. (Eds.), Course in General Linguistics. McGraw Hill, New York (Wade Baskin, Trans.).
Shah, R.R., Yu, Y., Shaikh, A.D., Tang, S., Zimmermann, R., 2014. ATLAS: automatic temporal segmentation and annotation of lecture videos based on modelling transition time. In: Proceedings of the ACM International Conference on Multimedia. ACM, pp. 209–212.
Shannon, C.E., 1948. A mathematical theory of communication. Bell Sys. Tech. J. 27, 379–423, 623–656.
Shannon, C., 1956. The bandwagon. IRE Trans. Inf. Theory 3.
Soleymani, Mohammad Reza, Khandani, Amir Keyvan, 1991. Vector trellis quantization for noisy channels. In: Advances in Speech Coding. Springer, pp. 267–276.
Tajima, K., Port, R.F., Dalby, J., 1997. Effects of temporal correction on intelligibility of foreign-accented English. J. Phon. 25, 1–24.
Thompson-Schill, S.L., Ramscar, M., Chrysikou, E.G., 2009. Cognition without control when a little frontal lobe goes a long way. Curr. Dir. Psychol. Sci. 18 (5), 259–263.
Tomasello, M., 2008. Origins of Human Communication. MIT Press, Cambridge.
Warren, R., Obusek, C., 1971. Speech perception and phonemic restorations. Percept. Psychophys. 9, 358–363.
Wittgenstein, L., 1952. Philosophical Investigations. Blackwell, London.
Wray, A. (2013) Why are we so sure what a word is? Mspt for John R. Taylor (ed.) Oxford Handbook of the Word, Oxford U. P. To appear, 2015, http://dx.doi.org/10.1093/oxfordhb/9780199641604.013.032.
Zipf, G.K., 1949. Human Behavior and the Principle of Least-effort. Addison-Wesley, Cambridge, MA.