



Production, comprehension, and synthesis: a communicative perspective on language

Michael Ramscar* and Harald Baayen

Department of Linguistics, University of Tübingen, Tübingen, Germany

*Correspondence: michael.ramscar@uni-tuebingen.de

Edited by:

Charles Jr. Clifton, University of Massachusetts Amherst, USA

Reviewed by:

Charles Jr. Clifton, University of Massachusetts Amherst, USA

MacDonald (2013) presents a strong case for speech production having a pivotal role in the cognition of language processing. Experimental research has been strongly biased toward the study of language comprehension, and it is an intellectual pleasure to be invited to rethink the consequences of the constraints imposed by speech production on both the form of utterances and how utterances are produced and understood.

Yet, it seems to us that production and comprehension are much more in balance. For instance, when Latin lost many of its inflectional exponents and morphed into what is now modern French, the pronouns of Latin, which were used for emphasis only, became obligatory. This, it would seem, serves the listener rather than making life easier for the speaker. In the convection cycle of language change over time, speakers time and again opt for articulatory simpler forms whenever they can. In French, this led to reduced forms (compared to Latin) of the subject and object pronouns. In modern colloquial French, these pronouns can even become prefixoids that are fusing into the verb, leading to structures such as *Jean il l'a vue Pierre*. The result is an inflectional system with subject and object marking on the verb, remarkably similar to the forms of Amerindian languages (Vendryes, 1921; Lambrecht, 1981). Simplification by the speaker is followed by diversification for the listener, which is followed by simplification by the speaker. Crucially, in the negotiation of communication, utterances only have a chance of being replicated (in the evolutionary sense) if they are both producible and understandable (cf. Steels, 1998; Steels and Wellens, 2006).

However, rather than attempting to evaluate MacDonald's program by

means of individual case studies, in this commentary we take a step back, and argue for a view in which the forces of production and comprehension are not only much more balanced, but in which they are essentially the same. To understand why we think the similarities are much more important than the differences, we turn to learning theory and information theory.

As MacDonald emphasizes, *learning* is a ubiquitous aspect of experience. Although, it is often conceptualized abstractly as a process that increases knowledge (like adding entries to an encyclopedia) and that improves performance (by increasing counters in the head, whether conceptualized as Bayesian priors or by serial search in a frequency ordered encyclopedia), it is important to note that the mechanistic picture of learning that has emerged from many lines of inquiry in the cognitive and brain sciences is *discriminative*. At both low- (e.g., O'Brien and Raymond, 2012) and high- (e.g., Ramscar et al., 2013b) levels of abstraction, learning is a process that reapportions attentional/representational resources in order to maximize future predictive success (e.g., Rescorla and Wagner, 1972; Pearce and Hall, 1980; Sutton and Barto, 1998; McLaren and Mackintosh, 2000; Schultz and Dickinson, 2000; Kruschke, 2001; Danks, 2003). Prediction error is used to *discriminate* against uninformative cues and to reinforce *informative* cues. These models of learning belong to a broad class of discriminative algorithms, along with the overwhelming majority of biologically based learning models (Schultz, 2006).

An important, though little-mentioned feature of this kind of learning is that it yields an inherently lossy form of coding (Ramscar et al., 2010). If languages are

learned discriminatively, the representations of relationships between form and meaning that learners acquire from experience will be subject to constant change, and these changes will involve information loss. Learned relationships between forms and meanings will be subject to constant variation, both across different language users, and within language users over time (Ramscar et al., 2013d). As MacDonald rightly observes, in these circumstances, *all* linguistic communication can be expected to involve ambiguity.

A crucial consequence of lossy coding is that linguistic forms do not simply serve as hash codes for mapping form onto meaning. The forms of language are simply not rich enough data structures to formally encode the full richness of the experiences they serve to communicate (Ramscar et al., 2010). It is therefore not at all clear what it means to say, as MacDonald does, that "linguistic utterances clearly differ from other actions in that they have both a goal (e.g., to communicate) and a meaning." Given what we understand about learning and encoding (see Grünwald and Vitányi, 2003 for an introduction to coding theory), it is clear that utterances neither *encrypt* their meanings, nor do they map onto them in a compositional, or even determinate, way. In spite of the pervasiveness of the structural metaphor (Lakoff and Johnson, 1980) that language is like a conveyor belt transporting boxes with meanings from speaker to listener, and that it is desirable to optimally stack the boxes so that their load is uniformly distributed over the conveyor belt (Hale, 2006; Levy, 2008; Jaeger, 2010; see Ferrer-i-Cancho and Moscoso del Prado Martín, 2011; Pellegrino et al., 2011 for critiques) there is good reason to believe that meaning is not *in* the words nor *in* the sentences.

This is where Shannon (1948)'s mathematical theory of communication provides insight:

“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. 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. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.” (Our emphasis.)

In other words, whatever the experiences and goals we wish to communicate might be, a signal should not be assumed to be a compositional deconstruction of them. Instead, an encoding simply needs to enable senders and receivers to discriminate between experiences and goals *on the basis of a shared code*. For example, in a world with just two experiences (being hungry; being satiated) and no noise, a code with just two non-decompositional signals, 0 and 1, suffices.

The relationship between signals and meanings in this kind of system can be summarized as follows (MacKay, 2003):

1. A communication *system* requires a sender and a receiver to be in possession of a source code defining the scope of the possible messages that can be transmitted.
2. Communication across the system is not concerned with the *meaning* of messages. In a Shannon system the receiver *reconstructs* the source message from the received signal by *discriminating* the source message from other possible messages that might have been selected and noise introduced by the communication channel.
3. The receiver does not interpret or expand on the source message. It simply reconstructs it at the destination with no loss of signal content. In linguistic terms, necessary condition for successful communication is that a

listener be able to correctly identify the form of the message sent. To the extent that a speaker and listener's codes converge, this will serve to reduce, or even eliminate, a listener's uncertainty about the experiences and goals that led a speaker to select that message, aligning the listener's predictions with the speaker's intentions.

Although, this picture is *very* different to most historical approaches to language (Frege, 1892; Russell, 1905; Wittgenstein, 1947; Miller, 1951; Chomsky, 1957, 1997; Tomasello, 2005), there are many reasons to believe that Shannon's theory provides a fruitful framework for the understanding of human communication.

First, as we noted above, learning is a process that leads to the acquisition of exactly the kind of predictive, discriminative codes that information theory specifies for artificial systems (Hentschel and Barlow, 1991; Atick, 1992). The critical difference between human and artificial communication systems is that human communicators *learn* as they go. Indeed, an alternative description of the goal of utterances is that speakers intend listeners to learn something from them. Virtually all utterances—even, “Hello!”—are intended to reduce a listener's uncertainty, whether about the world, or the thoughts, feelings etc., of a speaker; learning is largely defined in terms of this kind of uncertainty reduction (Rescorla, 1988; Hentschel and Barlow, 1991; Ramskar et al., 2013b).

Second, since learning is a discriminative process, acquiring a language amounts to learning how forms discriminate between the rich experiences and goals that speakers and listeners share (see Baayen et al., 2011, for a proof of concept). From this perspective, MacDonald's suggestion that prediction serves to “guide comprehension,”—somehow helping rich semantic understandings to be mysteriously extracted from a few sparse signals (Ramskar, 2010)—is unnecessarily vague and complicated when compared to a more straightforward view of comprehension as the reduction of listeners' uncertainty about speakers' intentions as messages unfold (Ramskar et al., 2010; see also Pickering and Garrod, 2007; McMurray and Jongman, 2011).

Third, not only does learning appear to extract a particular kind of predictive code (Schultz and Dickinson, 2000), but the distributional structures of languages correspond closely to *optimal predictive codes* (Hentschel and Barlow, 1991). In Shannon entropy terms, the least efficient possible code has a uniform distribution (i.e., one in which all alternatives are equiprobable at any given choice point) and the most efficient code is one in which items are distributed in the most non-uniform way possible (i.e., a power law distribution). The distributions of languages approximate the latter at every level so far examined (Zipf, 1949; Genzel and Charniak, 2002, 2003; Aylett and Turk, 2004, 2006; Manin, 2006; Futrell and Ramskar, 2011; Ramskar and Futrell, 2011; Piantadosi et al., 2011).

Finally, it is clear that the nature of learning changes across childhood (Ramskar and Gitcho, 2007; Thompson-Schill et al., 2009; Ramskar et al., 2013c). Very young children are deficient in many prefrontal functions that, as MacDonald emphasizes, are important to speech planning. This is a curious adaptation, but it offers at least one benefit: if “simple” discriminative learners are exposed to a highly structured environmental stimulus—a language and its experiential correlates—and are restricted to sampling it in the same, non-deliberative way, they will learn very similar *systems* of mappings (Ramskar et al., 2013a; see also Shannon, 1956).

In other words, learning, and its developmental trajectory across childhood, are particularly well-adapted for the acquisition of *common* predictive codes (in the Shannon sense), and linguistic distributions appear to have evolved—socially—to optimize these codes for communication (in the Shannon sense). It is within this information-theoretic rethinking of language that the question of the relative importance of comprehension and production in shaping language comes to stand in a different light.

We immediately acknowledge that linguistic distributions must be optimized for speech production (see also Zipf, 1949). However, we contend that this optimization is totally constrained by what the listener can tolerate. For instance, in spoken Dutch, the word *eigenlijk* (actually) can

reduce to *egk*. However, the speaker cannot opt for articulatory laziness in total disregard of the listener. Native speakers of Dutch do not understand *egk* when spoken in isolation (Ernestus et al., 2002; Kemps et al., 2004), and successful comprehension critically depends on its use in appropriate contexts. In other words, *egk* is a functional element of the speech signal by the grace of being part of a code that speakers and listeners share. Thanks to this shared code, what is easy for the speaker to produce is easy for the listener to understand. Likewise, what is more difficult for the speaker to encode, at whatever level of linguistic structure, is more difficult for the listener to decode. These considerations lead to the prediction that for each of the interesting examples discussed by MacDonald where we currently see optimization for production at work, there is a corresponding benefit for comprehension. If, as we suspect, Shannon's view of communication is correct, these benefits *must* be there, even if it is difficult to discern them at present, given our still very limited understanding of the experiences, and their neuro-cognitive instantiations, that we share when communicating with language.

ACKNOWLEDGMENTS

This research was made possible by an Alexander von Humboldt award to the second author.

REFERENCES

- Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? *Network* 3, 213–251.
- Aylett, M. P., and Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang. Speech* 47, 31–56.
- Aylett, M. P., and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *J. Acoust. Soc. Am.* 119, 3048–3058.
- Baayen, R. H., Milin, P., Durdevic, D. F., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychol. Rev.* 118, 438–481.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Chomsky, N. (1997). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.
- Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *J. Math. Psychol.* 47, 109–121.
- Ernestus, M., Baayen, R. H., and Schreuder, R. (2002). The recognition of reduced word forms. *Brain Lang.* 81, 162–173.
- Ferrer-i-Cancho, R., and Moscoso del Prado Martín, F. (2011). Information content versus word length in random typing. *J. Stat. Mech.* 2011:L12002. doi: 10.1088/1742-5468/2011/12
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und Philosophische Kritik* 100, 25–50.
- Futrell, R., and Ramskar, M. (2011). “German grammatical gender manages nominal entropy,” in *Presentation at Information-Theoretic Approaches to Linguistics 2011* (Columbus: LSA Linguistic Institute, Ohio State University).
- Genzel, D., and Charniak, E. (2002). “Entropy rate constancy in text,” in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)* (Ann Arbor, MI: Association for Computational Linguistics).
- Genzel, D., and Charniak, E. (2003). “Variation of entropy and parse tree of sentences as a function of the sentence number,” in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (Sapporo)*, 65–72.
- Grünwald, P. D., and Vitányi, P. M. (2003). Kolmogorov complexity and information theory. With an interpretation in terms of questions and answers. *J. Logic Lang. Inform.* 12, 497–529.
- Hale, J. (2006). Uncertainty about the rest of the sentence. *Cogn. Sci.* 30, 643–672.
- Hentschel, H. G., and Barlow, U. B. (1991). Minimum entropy coding with Hopfield networks. *Network* 2, 135–148.
- Jaeger, T. F. (2010). Redundancy and reduction: speakers manage syntactic information density. *Cogn. Psychol.* 61, 23–62.
- Kemps, R., Ernestus, M., Schreuder, R., and Baayen, R. H. (2004). Processing reduced word forms: the suffix restoration effect. *Brain Lang.* 90, 117–127.
- Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *J. Math. Psychol.* 45, 812–863.
- Lakoff, G. J., and Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago.
- Lambrecht, K. (1981). *Topic, Antitopic, and Verb Agreement in Non-Standard French*. Amsterdam: John Benjamins Publishing Company.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition* 106, 1126–1177.
- MacKay, D. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.
- Manin, D. (2006). Experiments on predictability of word in context and information rate in natural language. *J. Inform. Process.* 6, 229–236.
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Front. Psychol.* 4:226. doi: 10.3389/fpsyg.2013.00226
- McLaren, I. P. L., and Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Anim. Learn. Behav.* 28, 211–246.
- McMurray, B., and Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychol. Rev.* 118, 219–246.
- Miller, G. A. (1951). *Language and Communication*. New York, NY: McGraw-Hill.
- O'Brien, J. L., and Raymond, J. E. (2012). Learned predictiveness speeds visual processing. *Psychol. Sci.* 23, 359–363.
- Pearce, J. M., and Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* 87, 532–552.
- Pellegrino, F., Coupé, C., and Marsico, E. (2011). A cross-language perspective on speech information rate. *Language* 87, 539–558.
- Piantadosi, S. T., Tily, H., and Gibson, E. (2011). The communicative function of ambiguity in language. *Cognition* 122, 280–291.
- Pickering, M. J., and Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends Cogn. Sci.* 11, 105–110.
- Ramskar, M. (2010). Computing machinery and understanding. *Cogn. Sci.* 34, 966–971.
- Ramskar, M., Dye, M., and McCauley, S. (2013a). Error and expectation in language learning: the curious absence of “mouses” in adult speech. *Language*. (in press).
- Ramskar, M., Dye, M., and Klein, J. (2013b). Children value informativity over logic in word learning. *Psychol. Sci.* (in press).
- Ramskar, M., Dye, M., Gustafson, J. W., and Klein, J. (2013c). Dual routes to cognitive flexibility: learning and response conflict resolution in the dimensional change card sort task. *Child Dev.* (in press).
- Ramskar, M., Hendrix, P., and Baayen, R. H. (2013d). *Nonlinear Dynamics of Lifelong Learning: the Myth of Cognitive Decline*. Manuscript, University of Tuebingen.
- Ramskar, M., and Futrell, R. (2011). *The Predictive Function of Prenominal Adjectives*. Presentation at *Information-Theoretic Approaches to Linguistics 2011*. LSA Linguistic Institute, Ohio State University.
- Ramskar, M., and Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends Cogn. Sci.* 11, 274–279.
- Ramskar, M., Yarlett, D., Dye, M., Denny, K., and Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cogn. Sci.* 34, 909–957.
- Rescorla, R. A. (1988). Pavlovian conditioning: it's not what you think it is. *Am. Psychol.* 43, 151–160.
- Rescorla, R. A., and Wagner, A. R. (1972). “A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement,” in *Classical Conditioning II: Current Research and Theory*, eds A. H. Black and W. F. Prokasy (New York, NY: Appleton-Century-Crofts), 64–99.
- Russell, B. (1905). On denoting, in *Mind*, New Series, Vol. 14. Basil Blackwell, 479–493.
- Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol.* 57, 87–115.
- Schultz, W., and Dickinson, A. (2000). Neural coding of prediction errors. *Annu. Rev. Neurosci.* 23, 473–500.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423, 623–665.
- Shannon, C. E. (1956). The bandwagon. *IRE Trans. Inform. Theory* 2, 3.

- Steels, L. (1998). The origins of syntax in visually grounded robotic agents. *Artif. Intell.* 103, 133–156.
- Steels, L., and Wellens, P. (2006). “How grammar emerges to dampen combinatorial search in parsing,” in *Symbol Grounding and Beyond, Proceedings of the Third EELC*, eds P. Vogt, Y. Sugita, E. Tuci, and C. Nehaniv (Berlin: Springer Verlag), 76–88.
- Sutton, R., and Barto, A. G. (1998). *Reinforcement Learning*. Cambridge, MA: MIT Press.
- Thompson-Schill, S., Ramskar, M., and Chrysikou, E. (2009). Cognition without control: when a little frontal lobe goes a long way. *Curr. Dir. Psychol. Sci.* 18, 259–263.
- Tomasello, M. (2005). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Vendryes, J. (1921). *Le Langage*, Vol. 3. Paris: Albin Michel.
- Wittgenstein, L. (1947). *Tractatus Logico-Philosophicus*. New York, NY: Kegan Paul, Trench, Trubner and Company.
- Zipf, G. K. (1949). *Human Behavior and the Principle of Least-Effort*. Cambridge, MA: Addison-Wesley.
- Citation: Ramskar M and Baayen H (2013) Production, comprehension, and synthesis: a communicative perspective on language. Front. Psychol. 4:233. doi: 10.3389/fpsyg.2013.00233*
- This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology. Copyright © 2013 Ramskar and Baayen. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.*

Received: 12 February 2013; accepted: 11 April 2013; published online: 02 May 2013.