Challenges to developing time-based signal detection models for word production

Royce Anders, Alario F.-X.

Challenges to developing time-based signal detection models for word production

Royce Anders, F.-Xavier Alario

LPC UMR 7290, Aix-Marseille Université, CNRS, France

Note: pre-print version
Please check
https://doi.org/10.1080/02643294.2019.1581603

Keywords: language production, semantics, lexical selection, cognitive-behavioral modeling, response time analysis

Author note: Research supported by grants ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI) and the Excellence Initiative of Aix-Marseille University (A*MIDEX).

Corresponding author: Royce Anders, email: royce.anders@univ-amu.fr, address: UMR 7290 case D, Aix-Marseille Université & CNRS, 3 Place Victor Hugo, 13331 Marseille cedex 3

All Author Contact Details:

Royce Anders : royce.anders@univ-amu.fr

F.-Xavier Alario : francois-xavier.alario@univ-amu.fr

Challenges to developing time-based signal detection models for word production

A crucial process underlying language production is word selection (Levelt, 1989). In this process, the semantic components of an intended idea must link to the lexicon and activate the appropriate lexical item(s). Word selection, as driven by *lexico-semantic mapping,* must occur rapidly and precisely for language to be both fluent and well-aligned with a speaker's intentions. Lexico-semantic mapping is described by major psycholinguistic models (e.g. Dell, 1986; Roelofs, 1992) in terms of *spreading activation* dynamics. In such neural network models (NNMs) activation spreads through a layer of semantic nodes that sends inputs, via varying connection strengths, to a layer of lexical nodes (e.g. Chen & Mirman, 2012; Dell & Gordon, 2003; Oppenheim et al., 2010). While spreading activation is a dynamic that is well-agreed upon, Nozari & Hepner (2018) addressed a much less understood mechanism, the *selection rule*: how accumulating word activation levels are used to select a target for production.

Nozari & Hepner (2018) specify selection rule dynamics by extending a paradigm of conflict monitoring. They propose a *selection criterion*, that is the required distance needed between the accumulating activities of the highest and the next-highest word[1], which would trigger selection of the former. Hence the larger the distance between activations, the less internal conflict there is. They argue that the criterion is flexible, meaning that individuals may adjust their criterion to manage their internal conflict, which is introduced by task goals and/or speaking impairment(s). By adjusting one's criterion, fluency (speed) is traded for accuracy and vice versa.

At first this may seem too convenient, but trading speed for accuracy is actually a well-established feature of human performance (Pew, 1969). For a number of decades, this dynamic has been extensively modeled empirically, by a tradition known as evidence accumulation (see Ratcliff & McKoon, 2008; Voss & Voss, 2004, for reviews). The most popular model of this framework, the drift diffusion model (DDM, Ratcliff

[1] They also suggest that other activation differences could be explored.

1978), defines the criterion identically as in Nozari & Hepner's demonstration, but is not (yet) a model for word production. Other model variants, are more general in their definition of the criterion (Usher et al., 2002). In this work, we explore the connections between Nozari & Hepner's proposal and the framework of evidence accumulation.

What is evidence accumulation? Evidence accumulation constitutes a class of empirical cognitive-behavioral models that are used to fit performance data, and draw inferences about latent processes, especially selection dynamics. These models are known as an extension of signal detection theory to the time domain (Pike, 1973). Within a trial, evidence accumulation models activity accumulation (per millisecond) in preference for a response, toward a response-selection criterion. Since preference toward a response can be calculated in various ways, this mechanism can be directly linked to Nozari & Hepner's calculation of inverse conflict, as well as to other proposals (e.g. booster mechanism in Oppenheim et al., 2010).

Seen from this perspective, Nozari & Hepner (2018) contribute to a growing trend of using this type of framework to understand lexical selection and word production (e.g. Anders et al., 2015; 2017; Van Maanen & Van Rijn, 2007). But unlike Anders et al., instead of fitting specific data trends with accumulation models to inform NNMs, Nozari & Hepner come at this issue from a different perspective: they reduce NNM components to a framework of accumulation (of inverse conflict), and successfully account for some general data trends through that.

Given this growing trend of using traits from evidence accumulation to propose a selection mechanism in language production, how much further can we specify these dynamics, or support current proposals, without adequate empirical modeling? We discuss both the challenges and potential benefits to developing such empirical models for psycholinguistic and language production research.

Challenges to determining selection dynamics Selection criteria such as those based on absolute lexical activations or differential lexical activations (i.e. conflict) are not necessarily new to major psycholinguistic models, or NNMs (e.g. Dell & Gordon, 2003; Nozari et al., 2011; Oppenheim et al., 2010; Roelofs, 1992). But a principal roadblock for further specifying selection dynamics beyond current understandings (e.g. what types of activation computations are monitored, or how the selection criterion is determined), is often a problem of *model identifiability* or *mimicry*. For example, NNMs with different selection criteria perform equally well in reproducing mean trends of observed speech performance (see Simulations 5 & 6, Oppenheim et al.). Nozari & Hepner (2018) also relate their current model only to mean trends.

One way to discriminate between model mechanisms is to hold them accountable for more details in the data than just mean performance (i.e., additional data points, the variance, the distribution, single trials). Fewer models (with different mechanisms) will be able to fit equally-well the more complex features of the data, or at least multiple important data points than just one, especially by-participant. Based on this model mimicry problem, since Nozari & Hepner (2018) have not fit any data, we understand their predictions of how the criterion may move in impaired speakers as interesting hypotheses, yet to be assessed by formal model fitting comparisons with model-selection statistics.

In this context, the value of developing improved empirical modeling techniques is to provide a data-driven approach for evaluating cognitive dynamic proposals. For example in patients, Nozari & Hepner (2018) propose criterion adjustment is strategical in light of impairments, while Anders et al. (2017) propose an impairment in adjusting the criterion itself (we should acknowledge that we reasoned on correct responses only, see paper for motivation, as well as below). Without additional model fitting to new data sets (and improvements therein), it is difficult to tease apart that it is not both accounts at play. For example, is the conflict monitoring system also impaired in certain patient types, so the information those patients use to determine the criterion is very noisy? Could certain patients have cognitive control impairments in establishing and/or retaining a language criterion adjustment? Thirdly, what about conflict itself (differences between word activation levels), could there be noise in not only the word activation levels (and monitoring) but also in the difference calculation that is used to determine said conflict measure?

Can accumulation models alone resolve all of these questions? Such models still require developments themselves, but in applying them to our data sets, they are an added technique to explore these questions more quantitatively. The evidence accumulation approach provides an empirical modelling framework that is a standard above current fitting procedures; it can also fit observed data more closely than the mean trends considered in Nozari & Hepner (2018), as its procedures fit the entire distribution of response times (RTs); and they have the potential to quantify the selection mechanism they propose.[2]

Based on the cognitive accumulation parameters derived, these models can simulate an RT distribution that resembles the complexity of the observed data, and with a low residual. Therein, the mean, variance, and skew of the distribution are usually well recovered (e.g. see Anders et al., 2016), the residual across all observed-versus-predicted data points is often low, and each trial is subject to the "law" that an accumulation process therein occurred with the cognitive parameters estimated. Trial-by-trial predictions are also possible, in which an activation rate or criterion is estimated per every trial (Ratcliff et al., 2016). Though trial-by-trial parameters have not yet been modeled in language production, as it is difficult to estimate this many parameters when simultaneously, many experimental conditions are also modeled.

Current state of the art So far, accumulation models have been used to empirically model selection dynamics based on response times from conditions in blocked cyclic picture naming of both healthy speakers (Anders et al., 2015) and left frontal patients (Anders et al., 2017), as well as picture-word interference of healthy speakers (Anders et al., 2016, Experiment 2). The first findings with such an approach were respectively, a teasing apart of semantic interference and facilitation into separate parameters, criterion placement differences in left frontal patients, and semantic distractor effects on the criterion in picture-word interference.

A compatible accumulation model paradigm is the race framework (see Usher et al., 2002; Van Maanen & Van Rijn, 2007), in which each possible word can have an accumulator representing its preference or "inverse conflict." But explicitly, empirically modeling every word's inverse conflict involves an

[2] Note however that they are not exempt from their own issues of tackling mimicry (Donkin et al., 2011).

exponential number of trials needed. Anders et al. (2015) demonstrate a solution that by focusing on the accumulator of the observed word produced, current experiment sizes (e.g. with low repetition of the same word set) can be modeled empirically. Developments in big data may alleviate these limitations.

In current experiments, data limitations are still present even with the single accumulator case however. When modeling patients for the first time, Anders et al. (2017) did not aim to model the errors, since in their current setup, fifteen of the same error type per condition and per patient, were required to appropriately model them--numbers simply not present in the data. Nozari & Hepner (2018) rightfully claim that caution is warranted in how those patient modeling results are interpreted. Given the lack of sufficient empirical data modeling by both authors, additional fits to new data sets would be useful; and stability of parameters across experiments should also be assessed. Furthermore, modeling improvements are also a priority, such as finding a way to appropriately model error data despite the lack of appropriate (same) error observations in current experiments.

How could the Nozari & Hepner model be assessed by an empirical evidence accumulation application? Nozari & Hepner (2018) demonstrated a model of conflict monitoring between the target word and the next-most activated word. In the evidence accumulation tradition, this relates most directly to the drift diffusion model (DDM, Ratcliff, 1978; Ratcliff & McKoon, 2008). The DDM is an accumulator model like that in Anders et al. (2015) but with two boundaries, for example one associated with ”CAT”, and the other, ”DOG”. In the DDM however, activation is purely differential between the two words. That is, activation in favor of ”CAT” is 100% against activation of ”DOG”; this is numerically equivalent to Nozari & Hepner’s example of conflict (or its inverse). The DDM is appropriate for tasks limited to two choices, but the problem is word selection is a multi-choice paradigm, one for which race accumulation models are intended. The problem with using the DDM, or likewise modeling a direct interpretation of Nozari & Hepner’s demonstration, is that (*i*) throughout a trial (all milliseconds) only the same two words would be in contest, (*ii*) it could not handle conflict as determined from 3 or more words (and when word ranks change across milliseconds), and (*iii*) it would likewise assume that conflict is determined as a mere full subtraction (e.g. ”CAT” - ”DOG”), and only between these two words at each time point.

Despite their suggestion that other conflict measurements may be used, in their current proposal Nozari & Hepner provide no explicit instruction how the monitoring and criterion mechanism would handle three or more words, and would relate to the binary “correct” and “false” sensing distributions of conflict monitoring; moreover it is not clear how this could be explicitly modeled with an accumulation model. We look forward to such developments, such that an empirical modelling can be pursued in support of advancing their proposed, and valuable framework.

References

Anders, R., Alario, F.-X., & Van Maanen, L. (2016). The shifted Wald distribution for response time data analysis. *Psychological Methods*, *21*.

Anders, R., Rie`s, S., van Maanen, L., & Alario, F.-X. (2015). Evidence accumulation as a model for lexical selection. *Cognitive Psychology*, *82*, 57–73.

Anders, R., Rie`s, S., van Maanen, L., & Alario, F.-X. (2017). Lesions to the left lateral prefrontal cortex impair decision threshold adjustment for lexical selection. *Cognitive Neuropsychology*, (pp. 1–20).

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *119*, 417.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*, 283.

Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon: Friend or foe? In N. O. Schiller, & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 9–37). Walter de Gruyter.

Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E. J. (2011). Diffusion versus linear ballistic accumulation: different models but the same conclusions about psychological processes?. *Psychonomic Bulletin & Review, 18*(1), 61-69.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? a conflict-based account of monitoring in speech production. *Cognitive psychology*, *63*, 1–33.

Nozari, N., & Hepner, C. R. (2018). To select or to wait? the importance of criterion setting in debates of competitive lexical selection. *Cognitive Neuropsychology*, (pp. 3–40).

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*, 227–252.

Pew, R. W. (1969). The speed-accuracy operating characteristic. *Acta Psychologica*, *30*, 16–26.

Pike, R. (1973). Response latency models for signal detection. *Psychological Review*, *80*, 53.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: current issues and history. *Trends in Cognitive Sciences, 20*(4), 260-281.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142.

Usher, M., Olami, Z., & McClelland, J. L. (2002). Hick's law in a stochastic race model with speed–accuracy tradeoff. *Journal of Mathematical Psychology*, *46*, 704–715.

Van Maanen, L., & Van Rijn, H. (2007). An accumulator model of semantic interference. *Cognitive Systems Research*, *8*, 174–181.

Voss, R. K., A., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*, 1206–1220.