A brief survey of the computational hypotheses used to model language production

F.-Xavier Alario

CNRS & Université de Provence (UMR 6146)
Marseille

F.-Xavier ALARIO
Laboratoire de Psychologie Cognitive
CNRS & Université de Provence
3, place Victor Hugo - Case 66
13331 Marseille Cedex 3, France

+33 4 91 10 67 91
alario@up.univ-mrs.fr

Abstract

We review the computational hypothesis that have been used to model the psycholinguistic processes that subtend language production. The chapter begins with a general description of the architecture of the artificial neural networks that are currently used in this field. This section describes the basic hypothesis shared by most models (e.g., the structure of the network, the information coding scheme, the spreading activation equations). The second section focuses on the process of lexical selection – the process by which the word that is most appropriate for expressing a given message is selected. This section begins with a brief description of the psychological properties of the process that needs to be modeled. It is followed by an analysis of the spread of activation that is postulated in different models, a description of the selection criteria applied to the output units and, finally, a discussion of the modeling of neuropsychological lesions that affect the process of lexical selection. We reach the conclusion that, at this stage of the research in language production, computational models are better conceived as tools for hypothesis exploration than as faithful representations of the complete psychological process.

In this chapter we will describe some of the computational hypothesis that have been used to construct implemented models of language production. In other words, we are interested by the way in which cognitive hypothesis have been translated into mathematical form in order to construct the models. The purpose of the chapter is to provide an introductory descriptive overview of this issue. We will not attempt to formulate a new theoretical proposal, nor to evaluate the relative merits of the different models with respect to the empirical data they are intended to account for.

Within the field of psycholinguistics, research in language production investigates the processes that allow the retrieval of linguistic information from memory in order to express messages. An almost defining feature of this research is the fact that the linguistic processes under scrutiny are controlled by semantic or conceptual information (Levelt, 1989). This definition excludes a number of situations where language is in fact produced, for instance word reading. In this type of behavior, processing is under control of linguistic representations (i.e., the orthographic representations) rather than semantic or conceptual information. Although reading and other language processes have been extensively modeled (e.g., Ans, Carbonnel, & Valdois, 1998; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Plaut, McClelland, Seidenberg, & Patterson, 1996; Zorzi, Houghton, & Butterworth, 1998), we will limit our discussion to the implementation hypothesis used in models of language production[1]. The reasons for this are twofold. First, in recent years language production research has become a specific research field within the discipline of psycholinguistics (Wheeldon, 2000). Second, there are not many models of language production as defined in the previous strict definition. This makes the analysis we are interested in possible within the limits of this chapter.

The computational models that are used in language production research are complex and diverse. Furthermore, they do not necessarily share many features: they are constructed to account for different types of data sets (e.g., error patterns vs. reaction times), they use different methods (e.g., modeling simple global effects vs. complete distributions of data), they represent information by using different types of hypothesis (e.g., localist vs. distributed), etc. In this context, an analysis of the specific computational hypothesis that are made by the different models is useful for several reasons.

[1] Obviously, many of the modeling principles used to account for reading processes are used to account for language production processes (and vice-versa).

First, this analysis can help deconstructing the available models and their mechanisms. This should help determining the detailed reasons why a given effect can, or cannot, be simulated. Secondly, a better comprehension of the mathematical formulation should help checking its appropriateness as a reflection of the cognitive process it is intended to represent. Finally, this analysis is an essential step in the comparison of models that make use of different formalisms. As will become clear below, a great variety of implementation options have been taken by different authors.

We will consider those models of language production that have been proposed to account for single word production. In the remainder of this chapter, we will begin by presenting a few general principles of computational modelling. Then we will review the main implementations that have been used to account for lexical selection – i.e., the choice of the most appropriate word to express a given message. Lexical selection is the central process of many models of language production. For reasons of space, we will not consider other important aspects of production, such as phonological encoding (e.g., Dell, Burger, & Svec, 1997; Dell & O'Seaghdha, 1994) or the construction of syntactic structures.

1) The building blocks of the models

1.1) The structure of the neural network

In the vast majority of cases, the computational models of word production are based on groups of representational units. From a mathematical perspective, these units are numerical series whose index is a dimension that represents the (discrete) time in which the model evolves. These untis represent the information that is relevant for completing the behavior – for instance linguistic information, but also information that is not readily interpretable psychologically. The values taken by the units during the time sequence stand for the level of activation of the information that the unit represents. These levels of activation are bound by a system of equations that define their evolution as a function of the temporal variable. In the most general case, the activation value of a given unit is a combination of the values taken by other units at a previous moment in time, plus possibly other parameters. The equations that bind the activation of the different units represent the cognitive hypotheses being modeled. On the basis of this equation system and a set of "initial values", the activation of information can be defined for every moment in time (see Figure 1).

< Figure 1 here >

The activation values allow defining the responses "produced" by the model. Most of the time, this definition is based on further hypotheses that specify in one way or another how the "selection" is achieved (e.g., what are the relevant units? under which conditions do they need to be considered?), etc.) According to the selection criteria that is implemented, different types of language production data will be modeled (production errors or response latencies).

Some of the models are defined as deterministic functions whereas others are defined as random variables. In the first case, an initial state always leads, via the same trajectory, to the same final state. In the second case, the introduction of random variables provides, for each initial state, a probabilistic distribution of final states. From the psycholinguistic point of view, the second type of models opens the possibility of modeling inter-individual or inter-item variations by simulating the complete distribution of responses rather than a single average value.

1.2) Some characteristics of the representations

The representation of information in the model can follow a localist scheme (e.g., Dell, 1986) or a distributed scheme (e.g., Lambon Ralph, McClelland, Patterson, Galton, & Hodges, 2001). The localist implementation ascribes to each unit – or at least to a majority of the units – the activation value of a linguistic of psychological unit information. For example, the values taken by one of the units of the network (say unit B13) might represent the activation level of the lexical entry for the word "difference" (Page, 2000). In a distributed implementation, the activation of linguistic information – for example the lexical entry "difference" – is modeled by the levels of activation of a set of units in the network. These units will also code other linguistic information. For instance, the units labeled B01, B02, B03, ... to B20 represent (together) the set of words that the system "knows". Each known word (e.g., "difference") is coded by a set of twenty predefined activation values that the units can take. At any time *t*, the closeness (in terms of a pre-defined distance) of the vector [B01(t) …

B20(t)] to the pre-specified values for the lexical entry "difference" will index the level of activation of this unit in the memory of the system (Hinton, McClelland, & Rumelhart, 1986).

With few exceptions (the lexical selection models of Lambon Ralph et al., 2001, or Plaut & Shallice, 1993 ; the phonological encoding model of Dell, Juliano, & Govindjee, 1993), modeling of language production processes has been based on localist representations. A possible reason for that is that the information coding scheme in distributed models is less transparent than in localist models. Furthermore, the benefits that often motivate the use of distributed representations (e.g., learning, generalization, partial information inferences, etc.) are not those that are pursued in language production research.[2]

Localist and distributed models implement psychological hypothesis in rather different ways. One can compare these two families by considering localist representations as distributed representations that are strongly biased (Figure 2, left). In a localist model, the lexical entry for "difference" (for example) will generally be thought to be activated if unit B13 has an activation value that is much larger than the activation value of the units representing other lexical entries (the details of this statement depend on the selection criteria that is used). An interpretation of this coding scheme in distributed terms is possible if the coding or the lexical entry "difference" is defined as a vector with very low values in all of its coordinates, except for the 13th coordinate. This constraint on the representation of information is not present in distributed models where the coordinate values that represent the information can freely take the full range of values (see Figure 2, left). Other constraints also apply to localist models. Localist models postulate one representational unit for each unit of information to be represented, whereas distributed models generally postulate many more series (units) than psychologically relevant information bits. Also, localist representations are most often associated with restrictions on the connectivity in the model (see below). Such restrictions do not always apply *a priori* to distributed models.

< Figure 2 here >

[2] Dell et al. (1993) have used distributed representations of phonological information to show that this encoding does not require the structural representations that are often thought to be unavoidable at this level.

In short, localist representations are equivalent to distributed representations plus a number of restrictions on the interpretation of the numeric series and the information coding. Conversely, it is possible in certain cases that a distributed model is reformulated in localist terms, if the information structure allows it.

1.3) The models' connectivity

A model is composed of a set of units whose activation values generally depend of the activation values of other units in the system. Two units whose values depend on one another are said to be connected: activation can spread between the information they represent. If the relationship between the activation of the units is an increasing function, the connection is activatory; if the relationship between the activation of the units is a decreasing function, the connection is inhibitory. The general spreading activation formula currently used in most models of language is production is classic:

$$\text{activation}_i(t) = [\ \text{activation}_i(t\text{-}1) \times \text{d}\ ] + \Sigma_j\ [\text{activation}_j(t\text{-}1) \times \text{p}_{ij}] + \text{noise}$$

where *i:* current unit index

j: an index for all units connected to unit i

d: activation decay parameter

p_{ij}: connection weight between units i and j

t: discrete time

As can be seen, the activation of a given unit depends on its own activation and on the activation of the units it is connected to. An important parameter of the model is the relative weight of the two contributions to the activation of the units. This parameter can be made explicit by considering the contribution due to the unit itself (d) and the contribution due to the activation of other units (p x m, where p is the average connection weight and m the average number of units that reach a given unit). A unit's activation will be reluctant to change (it will have more "inertia") if $d > (p \times m)$. Conversely, a unit will be more sensitive to its context in the opposite case. Also, under most hypotheses, if $d + (p \times m) > 1$ then the units' activation will grow indefinitely over time whereas if $d + (p \times m) < 1$ they will converge to zero after a number of time steps. Exploring the values of these ratios in the different "levels" of a numerically specified computational model should help understanding its temporal evolution. As a matter of fact, the investigation of lesioned models relies partly on the investigation of how the modification of these parameters affects the behavior of the model (see section below).

Some variants of the equation given above are of course possible. These mostly concern (a) the possible use of non-linear functions to modulate the accumulation of activation and its limits (e.g., the sigmoid function used to bound activation within finite limits; Starreveld & La Heij, 1996), and (b) what units can contribute to the activation of a given unit (the connectivity). The details of the connectivity hypothesis allows defining families of units that have similar connectivity patterns and hence represent similar linguistic information. Levels of processing are in fact implemented in this way. In many language production models, the units within one such level are not inter-connected: they only have connections to the units of other levels (except in some cases, such as the semantic levels of the models in Harley, 1993; Laine, Tikkala, & Juhola, 1998; or Lambon Ralph et al., 2001). This absence of "lateral" connections" (i.e., connections within a given level) contrast with the hypotheses that are usually made when modeling other domains of language (e.g., word reading). In these other models, inhibitory lateral connections are often postulated. The computational role of these connections is to enhance the difference of activation between the units of a given level. In a model with lateral inhibitory connections, the unit that has the largest activation will increase the difference with respect to the other units in the course of time. Without these connections, the differences between units will not be amplified. In fact, the choice of using or not using lateral connections is closely tied to choice of a selection process. Selection criteria can require wide activation differences between the units; alternatively they might be effective on the basis of limited differences. As we will review below, language production models put a strong emphasis on the definition of the selection mechanism.

2) The lexical selection process

In this section, we review the hypotheses that specifically characterize the process of lexical selection. This process is responsible for the choice of the appropriate word among all the words that are known by the speaker.

2.1) The spread of activation: from serial to interactive

In a computational model, the activation spreads from a representational unit r1 to a representational unit r2 when the activation value of r2 depends on the activation value of r1. The activation of r1 at a given point in time generally contributes to the activation value of r2 at the next point in time. How should the units be connected in a model of lexical access? The issue of connectivity is one of the central questions addressed in the psychology of language. To provide a description of how this question is framed within the field of language production, we begin with some rather uncontroversial assumptions. In a given production situation (for example, in the experimental situation where a speaker has to name a picture), the language production system will first activate the semantic information that corresponds to the message that has to be expressed. This information will activate several lexical candidates among which the most appropriate will be selected. The system will also activate the relevant phonological properties. Within this context, the issue of the spread of activation in the language production system has been framed in the following way. (1) Do the lexical entries that are activated but that will not ultimately be produced activate also their phonological properties? In case the answer is "yes", the spread of activation will be described as "cascaded"; alternatively, the spread of activation will be "sequential". (2) Under the assumption that activation spreads in a cascaded fashion, does the pre-activated phonological information influence the lexical selection process upstream? If it can, the system is said to be interactive, otherwise it is said to be serial (see Rapp & Goldrick, 2000, for extensive discussion).

Most of the logically possible answers to these questions have been defended on empirical grounds: strict seriality (Levelt, Roelofs, & Meyer, 1999), cascaded processing without interactivity (Humphreys & Riddoch, 1988), bidirectional interactivity (Dell, Schwartz, Martin, & Gagnon, 1997), etc. Each one of these articles proposed a model with different characteristics that was compatible with a set of experimental data. However, these models did not always attempt to systematically evaluate whether minor variants of the models also accounted for the data. This is an important point, since Harley (1993) has been able to show that data sets that were used to favor a sequential model could easily be accounted for in a cascaded model. In fact, rather that seeking for a model that was able to account for a given data set, this author attempted to investigate the general properties of a class of computational hypothesis.

Following a similar logic, Rapp and Goldrick (2000) conducted an extensive evaluation of different spreading activation hypothesis (see also Dell, Martin, Saffran, Schwartz, & Gagnon, 2000; Dell, Schwartz et al., 1997; Foygel & Dell, 2000; Harley, 1993; Ruml & Caramazza, 2000; Ruml, Caramazza, Shelton, & Chialant, 2000). They re-formulated the dichotomy between seriality and interactivity as a continuum defined by several dimensions that characterize how information can be exchanged between representations: the connection weight that modulates the exchange of activation between units, the direction (forward-serial or backward-interactive) of this exchange, the level of the model where the exchange occurs, and finally, the amount of noise present in the system. They also manipulate the impact that a selection has on the level of activation of the unit that has been selected. This dimension also constrains the degree of seriality of the process. This is because a selection that has a high impact on the level of activation (e.g., if the selection produces an increase of activation of about ten times the current activation rather than, say, around 25% more of the current activation) will produce a discontinuity in the levels of activation of the units that are downstream of it. This discontinuity can be interpreted as a transition from one mode of processing (e.g., lexical processing) to a subsequent mode (e.g., phonological processing). Clearly, a stronger selection is equivalent to a more sequential processing (Dell & O'Seaghdha, 1991).

Overall then, this combination of manipulations allows defining a spread of activation that will be more or less cascaded, and that will involve more or less interactivity between the levels of the model. This approach to the problem of language production modeling illustrates how different computational hypothesis that are not mutually equivalent can be used to implement a given psychological hypothesis (namely, the degree of interactivity). The parametric approach allows defining a set of models that vary on the degree of interactivity they allow. The choice of the better model is then made on the basis of an optimization of the parameters of the continuum. Rapp et Goldrick's (2000) study indicates both what degree of interactivity is required to account for a given data set and why other degrees of interactivity (e.g., strict seriality) do not fit the data as well. An exhaustive exploration of the space of possible models given by a set of computational hypothesis is a key to a better understanding of why the hypothesis have to be implemented.

A clear advantage of the approach adopted by Rapp et Goldrick (2000) is that the space of possible models is explicitly specified and that it is completely explored. A possible limitation that is inherent to this approach is, of course, that the search for an appropriate model is limited *a priori* to that space. The approach excludes the possibility of testing other mechanisms (e.g., a monitoring mechanism in the case of lexical selection) not present in the original proposal and that could be better off at accounting for the data set being modeled.

2.2) Unit selection criteria

The unit's activation evolves over time on the basis of the connectivity defined in the model. Then some units have to be selected to form the response produced by the model. Selection criteria have been implemented in many different ways in models of language production. The criteria depend on the kind of data that are modelled and on the structure of the numerical series. When modelling production errors, a given initial state can lead to a variety of final states. One of them is considered to be correct whereas the others will be errors of different types. When modelling naming latencies, a given initial state can be associated to the number of cycles that need to be completed by the model to reach a specific final state.

One type of response criterion hypothesizes that the response of the model occurs at a pre-specified point in time. After a fixed number *n* of cycles the selected unit is the one that is the most active (localist model; Figure 3A), or the global pattern that is selected is the one observed at that point in time (distributed model). This implementation of the selection criteria is very simple, and indeed one of the most frequently used (e.g., Dell, 1986; Dell, Schwartz et al., 1997; Foygel & Dell, 2000; Rapp & Goldrick, 2000; Ruml et al., 2000; Vousden, Brown, & Harley, 2000). Given that the delay for selection is fixed, this selection criterion allows modeling the correctness of the responses but not their delay.[3] Furthermore, most of these models do not consider the actual level of activation reached by the units at the point in time where selection occurs (see below for exceptions). A model that implements selection by means of a fixed temporal criterion will provide correct results if the trajectories stemming from the different initial states are ordered appropriately – i.e., the most expected response being the most active, etc. – after the predetermined number of temporal cycles. In other words, the trajectories stemming from the initial states cross the temporal threshold in the right order.

A hypothesis that comes close to the fixed temporal criterion is the one that uses an absolute activation threshold. In this case, a unit will be selected if its activation exceeds a fixed value, irrespective of the value taken by the other units at that time (MacKay, 1987; Santiago, MacKay, Palma, & Rho, 2000). This type of implementation allows in principle the modeling of two characteristics of the responses: their latencies – the cycle of the model where the threshold is reached – and their correctness – what numerical series reaches the threshold. This is because, contrary to the temporal threshold selection criterion, the activation of the unit is considered in this criteria and the interpretation of "latency" can be given to the moment in time where the threshold is reached. As was noted for the models with temporal criterion, a model with activation criterion will produce appropriate results if the activation functions reach the selection criterion in the proper order (figure 3B).

< Figure 3 here >

In fact, under certain circumstances, these two selection criteria are formally comparable, even equivalent. This is the case if the output units (whether they are considered individually or in sets) follow increasing functions (a relatively common situation). If this condition is met, and given a set of activation trajectories in time, selecting on the basis of a temporal criterion or on the basis of an activation criterion is equivalent, even if the trajectories of the units cross each other. In other words, a set of trajectories that is ordered at a given point in time will always be similarly ordered for a given activation level. Hence the model will be able to produce the same results with the two criteria. The

[3] The n value can be manipulated as a parameter, generally representing speech rate (a smaller n represents a faster speaker).

reason for this is that because increasing activation functions are bijective, hence the roles of time and activation can be inverted[4]. By contrast, if the activation functions are not increasing overall (Figure 3C) then the two selection criteria are not necessarily equivalent. The preference given to a temporal or activation criteria will depend on the detailed properties of the activation functions.

The selection criteria can be made more complex in order to account for other information present in the activation functions. It is possible to evaluate the global state of activation in the lexicon when selection occurs by considering the relative levels of activation between the functions. In a model with an activation based selection criterion, a standard method to do that is to consider the levels of activation of several units in the selection criterion. For example, a unit will be selected when its activation surpasses that of all the other units by a pre-specified amount of activation. (figure 3D ; Wheeldon & Monsell, 1994, discuss this algorithm but do not implement it; see also Starreveld & La Heij, 1996). The same approach can be applied to a model with a temporal threshold criterion. Dell, Burger et al. (1997) use the following rule: at the time *n* of selection, all the units stand a chance to be selected. This chance is proportional to their level of activation. In this way, the unit that has the highest probability of being selected is of course the most activated unit. Yet the system can produce errors, either because the most activated unit is not the expected target or because, even if it is, another unit ends up being selected.

An implementation that is somewhat similar to that one is that used by the influential and comprehensive model of lexical access WEAVER++ (Levelt et al., 1999; Roelofs, 1992, 1997). In WEAVER, the selection of the units is not based on criteria of activation or time. WEAVER implements so-called *verification* procedures that ensure that the unit selected at a given level is indeed the one that corresponds to the unit selected at the previous level (Roelofs, 2003). For instance, to account for the double input picture-word interference paradigm, WEAVER knows what unit it has to chose based on flags that mark the origin of the activation that reaches each unit[5]. A general

[4] The benefit of using activation based criteria is that they allow a direct psychological interpretation of the non-restricted dimension (time), provided of course that the obtained temporal patterns are appropriate.

[5] Marking the origin of activation might seem an unnatural hypothesis. This is only true if one imagines activation as an indiscriminate quantity, that is exchangeable without restrictions from one part of the model to

consequence of this is that the model is not equipped for the simulation of error patterns, since the responses that are produced are always verified and correct. WEAVER simulates latencies by considering the activation downstream of selection. In this model, the variability in the latencies that simulates the data patterns originates in the relative activation of the units. Once a unit is selected (on the basis of the external criteria of flagging), the time required to retrieve its properties is given by the ratio between the activation of the unit and the sum of the activation of all other relevant units. In other words, Dell, Burger et al. (1997) use relative activation levels to generate errors, whereas Roelofs (1997) uses them to generate response latencies.

In the different implementations we have discussed, the selection of information makes use of an external criteria: a number of cycles, an absolute or relative activation threshold, the labeling of numerical series, etc. This external criterion is required when the temporal evolution of the model does not provide an intrinsic definition of the responses. By contrast, in those models where the levels of activation converge on finite values, an intrinsic definition of the responses is possible. By definition, the variation of a convergent activation function is reduced over time. Postulating a criterion based on this variability[6] (e.g., a variability that does not exceed 1% of the value of the function) allows to simulate both the latency (the time required to reach stability) and the correctness (specific value reached) of the responses. This type of activation function is attractive, among other things, because it allows an intrinsic definition of the responses (an initial state produces a corresponding final state) and because it provides simulations of the two dimensions that are relevant for language production. To this day, this implementation of the selection criterion has only be used to model the correctness of the responses (e.g., Lambon Ralph et al., 2001), but not their latencies.

The consequences of selection are also an important parameter of the selection criterion. This issue was mentioned in the previous section on interactivity in models of language production. We

---

the other. However, if an experimental task is modelled and external processing hypothesis (e.g., attentional mechanisms) are introduced then the knowledge of the origin of the information could be modelled (see Roelofs, 2003, for extensive discussion).

[6] Just as the previous ones, this criteria is external. However, contrary to the previous ones, this criteria is not used to identify the output response. This response is given by the intrinsic value of convergence of the series. The criteria is only used to specify a required level of stability of this response.

noted that the degree of interactivity is to some extent depended on the consequences of selection at every level. The hypothesis that define the initial states of the system are also important. Several hypothesis are possible: the initial state can be represented by a constant fixed activation of the units representing the input, by a decreasing activation of these units or by a fixed activation maintained over a finite number of cycles. This simple difference can produce dramatically different patterns of activation at the level of the output units (figure 4).

One of the main challenges within the field of language production is the definition of a selection criterion that will be able to account, in a single model, for latency and error patterns. An ideal model will reproduce both the latency patterns and the different biases and regularities that are characteristic of erroneous productions. As a further challenge, the model should also be able to simulate anomic states (tip of the tongue in healthy speakers: Brown, 1991; or omissions in anomic patients: Laine et al., 1998), as well as other response types given by aphasic patients (e.g., circumlocutions: Ruml et al., 2000). An implementation that starts by specifying the requirements of the selection criterion (what kind of result the model should reach) could be a fruitful approach to construct a complete model.

2.3) Lesioning models and modeling lesions

One of the objectives of language production models is to account for the error patterns observed in healthy speakers and aphasics. In many cases, the modeling of errors is achieved in two steps. First, a model is constructed on the basis of various equations and parameters that essentially produce correct responses (sometimes these models have a marginal capacity to produce errors). Then, the numeric functions or their parameters are modified such that the equations do not longer produce the correct pattern, but rather the expected deviant patterns. Accordingly, error modeling offers two "degrees of freedom": in the initial definition of the model and in its subsequent modification.

The modeling of a lesion of the lexical system has to be implemented in relation with the neuropsychological hypothesis that is to be implemented. For example, it can be hypothesized that a lesion will affect in the same way all the levels of processing involved in a given task. This global lesion hypothesis has been successfully used in certain models of speech production (Dell, Schwartz et al., 1997; see discussion in Dell et al., 2000; Foygel & Dell, 2000), but its implementation remains controversial both on computational and neuropsychological grounds (Ruml & Caramazza, 2000). Alternative hypothesis are based on the idea that damage can affect selectively certain processes required for completing an experimental task, or that certain levels of processing are more affected by the lesion than others (see the discussion of the manipulation of the connectivity parameter below).

Also central for the production of errors is the contribution of noise to the levels of activation of the units. Noise is defined as a random variable, generally with a normal distribution around zero, whose variance depends on various parameters of the model. The first parameter is the severity of the deficit: a larger variance will be used for more severe deficits. A second parameter is whether the amount of noise is homogeneous across the units that are affected – so-called *intrinsic noise* – or whether it depends on the level of activation of the affected unit – so-called *activation noise*. Postulating that noise magnitude is homogeneous across affected units can produce undesired instability in the lexical system. This is because under this hypothesis a unit that is normally not very activated can receive the same amount of activation (or inhibition) than a unit that is normally very active (Rapp & Goldrick, 2000). As a consequence, the patterns of activation present in the "healthy" model can undergo considerable arbitrary changes. By contrast, if the magnitude of the noise is related to the level of activation, the units' trajectory deviations produced by the noise will preserve to some extent the pattern of activation present in the non-lesioned model. Since it is generally postulated that errors are constrained by the structures responsible for normal use (Fromkin, 1971), the second hypothesis seems preferable to account for lexical access in language production.

Besides manipulations of the magnitude and nature of the noise affecting the units' levels of activation, the parameters that define the normal functioning of the model can be modified in various ways in order to simulate a whole range of lesions (Dell, Schwartz et al., 1997; Foygel & Dell, 2000; Ruml et al., 2000). The first parameter that can be manipulated to simulate a lesion is the activation decay (d). In the general equation presented on page ##, this parameter modulates the amount of its own activation that a unit keeps from one time cycle to the next. The higher d is (the closest it is to 1), the more similarity there will be between the activation at a given cycle and at the next cycle. An approximate description, then, is that d indexes the temporal stability within the system. The modulation of *d* is one of the manipulations used by Dell, Schwartz et al. (1997) and by Foygel et Dell (2000) to create a space of possible models where they attempt to localize the aphasic patients they wish to simulate. Their simulation study shows that the modulation of *d* is in part responsible for variations in proportions of semantic and phonological errors (semantic and phonological errors are defined on the basis of their similarity to the expected target; non-word errors bear no-relationship with the expected target). They interpret this observation as evidence that the manipulation of *d* keeps the informational structure of the network intact: the opportunities for errors are constrained by the proximity in the original network (hence the occurrence of related errors). Rapp et Goldrick (2000) reach a similar conclusion when they argue that the manipulation of *d* reduces the level of activation overall present in the network and simply renders the effect of noise visible.

Somewhat different conclusions are reached by Ruml et al. (2000). Their manipulation of the parameter *d* leads to a very restricted set of possible models. The modulation induced by this parameter is very weak in the model they investigate and the models that are thus generated do not lead to very different error patterns. Consequently, they question the relevance of this manipulation. Also, a global analysis of the values taken by *d* in the fits of the models by Dell, Schwartz et al. (1997) or Ruml et al. (2000) shows that the variability in *d* is relatively weaker than the variability in *p* (the connection strength parameter). This suggests that the role of parameter *d* in adjusting the models to the patient's performance could be much smaller than that played by the connectivity parameter.

One explanation for these differing roles between parameters could be the context in which this manipulation is conducted. As we noted earlier, the equation that defines the spread of activation is composed of various terms. The importance of each of these terms for the global functioning of the model depends on their relative values. In a model where the value of p x m (mean connection strength X mean number of connections) is large with respect to *d*, the latter parameter should have a reduced role in the evolution of the activation of the units. Consequently, even if the manipulation of *d* alone provides information about the role of this parameter in the model, it cannot be substitute to an combined analysis of the different contributions to the spread of activation.

The other parameter that has been manipulated in order to simulate lesions in a model is *p*, the parameter that characterizes the strength of the transmission of activation between the units (connections). As we have seen, the definition of the connectivity is an essential element in the coding of information in the system. A manipulation of the connections directly affects the information that is present in the system. It can lead to a variety of outcomes depending on how the information is originally coded in the system. In certain models of lexical access all the connections have the same weight (the information is then coded by the presence or absence of link between the units). Under this hypothesis, lowering the global value of the connection strength parameter will affect the system globally. This manipulation is similar to limiting the amount of information that is coded in the connections. Dell, Schwartz et al. (1997) show that this manipulation leads to a reduced capacity of the model to produced "differentiated" levels of activation for the different units. This in turn leads to an increase of random errors (interpreted as non-word errors since they do not have a clear relationship to the intended target). In a variant of this manipulation, the connections are categorized and only some of them are manipulated (for example, those that connect two specific levels such as the semantic and phonological levels: Foygel & Dell, 2000). This lesion hypothesis is clearly more structured that the preceding one, and it integrates distinctions based on the neuropsychological theory and observations (e.g., the semantic – phonological distinction). Not surprisingly, it allows the simulation of much more specific deficits, such as the production of a high proportion of errors of a single kind (see also Rapp & Goldrick, 2000; Ruml et al., 2000).

In certain models, especially those that use distributed representations, the connection strengths are not all similar. Those models generally postulate a large number of connections whose values are fixed by a learning procedure that establishes the information coded in the system. Lesions are generally modeled by erasing some of the connections. For instance, 25% or 50% of the connections (chosen at random) between two levels can be set back to zero (Lambon Ralph et al., 2001). The impact of this type of lesion is difficult to evaluate *a priori* because of the distributed character of information encoding. In fact, only a detailed investigation of the lesions, especially one that takes into account the other parameters in the model, will characterize the impact of the lesion on the functioning of the model. Up to now this kind of investigation has been conducted only with numeric simulations, and an exact investigation might be useful

Before closing this section, we mention a relatively different method used to model the production of errors by a few studies. In this hypothesis, the noisy model produces errors with a pre-specified probability p_E. In the course of an experimental trial, the probability that the trial will be erroneous is p_E. If the trial is correct, then the model functions normally. If the trial is erroneous, the model functions normally except for the fact that the correct response cannot be selected; the actual response is selected among all the other candidates on the basis of their level of activation. Vousden et al. (2000) implement this system with a secondary constraint. The "distance" between the target unit and the unit selected by error should not be above a pre-specified value. In other words, the set of possible erroneous responses is constrained. In this implementation, the production of errors happens outside the language production system, since an independent random function determines whether or not an error has occurred. As such, the model is explicitly not a simulation of the actual mechanism that is responsible for the occurrence of errors in the speaker.

3) Conclusion

In this chapter we have reviewed some of the computational implementations of psychological hypothesis used to account for the processes of lexical access in language production. This brief survey has evidenced the variety of options that exist for implementing very close psychological hypothesis. This diversity complicates the comparison between models as well as the selection of the "most appropriate" or "most truthful" model. Hopefully, this current diversity should lead to a specification of the computational characteristics that are the most appropriate to obtain an exhaustive modeling of the process of lexical access. As a matter of fact, the existing models can be seen in two different ways: either as exact descriptions of the processes responsible for language production, or as tools for systematically exploring different possible descriptions of these processes. At this point, the latter option seems to remain as the unavoidable stage to get some day at the first one.

Figure captions

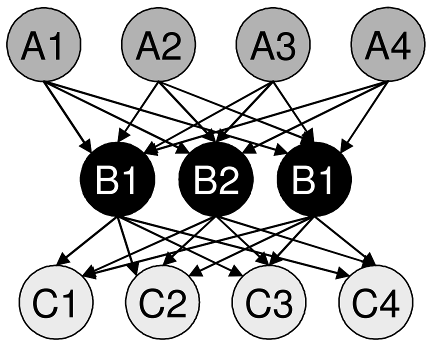
Figure 1:
Left: schematic representation of a baby-model of language production specified for illustrative purposes. This baby-model shares with many current models a structure in three levels, without connexions between levels and without connections between the first and the third levels (A → C). Right: an example of the levels of activation of the units (random parameters).

Figure 2:
Two different coding schemes for the lexical entry "difference" in a localist model (one of the units, B13, is responsible for coding this entry) and in a distributed model (the coding is based on a vector of values that characterizes all units).

Figure 3:
A- Temporal selection criterion: the selected unit (C4) is the most active at time $n = 10$
B- Activation selection criterion: the selected unit (C4) is the first to pass a pre-specified activation threshold.
C- Activation functions that do not allow establishing an equivalence between temporal and activation selection criteria; in this situation, the choice of the criterion will require a detailed evaluation of the activation functions.
D- Selection criteria based on a difference of activation: the selected unit (C4) is the one whose activation exceeds that of all the other units by a pre-specifed difference Δa_i.

Figure 4:
Activation trajectories in one of the output units of a model where the input activation is maintained during all the processing (I1), in a model where the input activation is constant over a finite period (I2), and in a model where the input activation decreases over a finite period (I3).

Figure 1



$$a_i(t) = 0.5 \; [1<t<10]$$
$$u_i(t) = u_i(t-1) * 0.6 + \Sigma \, [u_j(t-1) * p_{ij}]$$

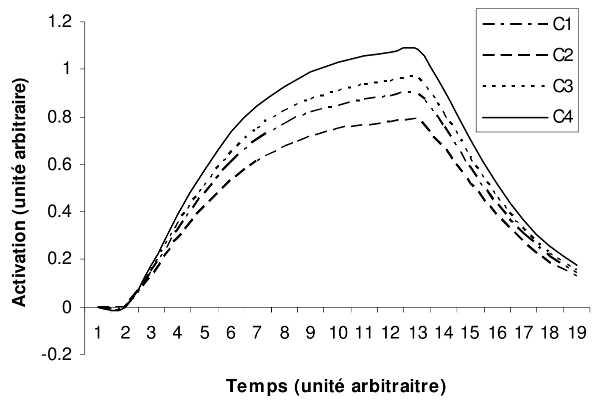


Figure 2

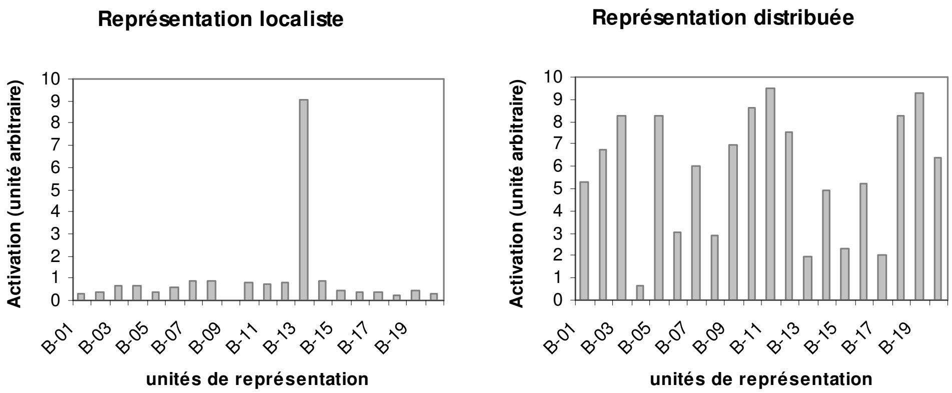

Représentation localiste
Activation (unité arbitraire)
10
9
8
7
6
5
4
3
2
1
0
B-01
B-03
B-05
B-07
B-09
B-11
B-13
B-15
B-17
B-19
unités de représentation
Représentation distribuée
Activation (unité arbitraire)
10
9
8
7
6
5
4
3
2
1
0
B-01
B-03
B-05
B-07
B-09
B-11
B-13
B-15
B-17
B-19
unités de représentation

Figure 3

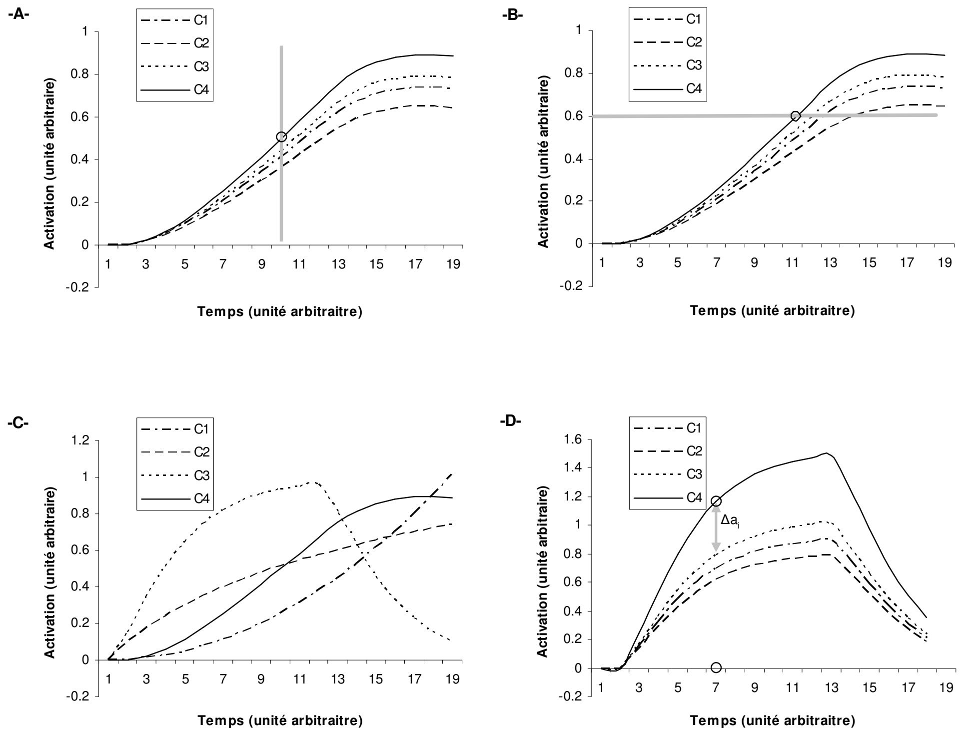
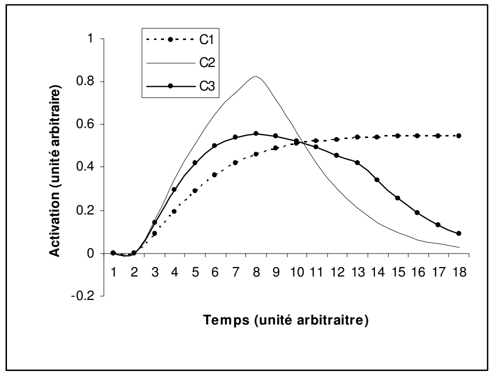

-A-
-B-
-C-
-D-
C1
C2
C3
C4
Activation (unité arbitraire)
Temps (unité arbitraitre)
Δa_i
1
0.8
0.6
0.4
0.2
0
-0.2
1.2
1.4
1.6
1 3 5 7 9 11 13 15 17 19

Figure 4


C1
C2
C3
1
0.8
0.6
0.4
0.2
0
-0.2
Activation (unité arbitraire)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Temps (unité arbitraitre)

REFERENCES

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review, 105*(4), 678-723.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109*, 204-223.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production in serial order: A functional analysis and a model. *Psychological Review, 104*, 123-144.

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production. *Cognitive Science, 17*, 149-195.

Dell, G. S., Martin, N. M., Saffran, E. M., Schwartz, M. F., & Gagnon, D. A. (2000). The role of computational models in neuropsychological investigations of language: Reply to Ruml and Caramazza (2000). *Psychological Review, 107*(3), 635-645.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al (1991). *Psychological Review, 98*, 604-614.

Dell, G. S., & O'Seaghdha, P. G. (1994). Inhibition in interactive activation models of linguistic selection and sequencing. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 409-453). San Diego, CA: Academic Press, Inc.

Dell, G. S., Schwartz, M. F., Martin, N. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*, 801-838.

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory & Language. Special Issue: Disorders of language and memory: Implications for cognitive theory, 43*(2), 182-216.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27-52.

Harley, T. A. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes, 8*, 291-309.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In the PDP research group (Ed.), *Parallel distributed processing*. Cambridge, MA: The MIT Press.

Humphreys, G. W., & Riddoch, M. J. (1988). Cascade processes in picture namig. *Cognitive Neuropsychology, 5*, 67-104.

Laine, M., Tikkala, A., & Juhola, M. (1998). Modelling anomia by the discrete two-stage word production architecture. *Journal of Neurolinguistics, 11*(3), 275-294.

Lambon Ralph, M. A., McClelland, J. L., Patterson, K., Galton, C. J., & Hodges, J. R. (2001). No right to speak? The relationship between object naming and semantic impairment:

Neuropsychological evidence and a computational model. *Journal of Cognitive Neuroscience, 13*(3), 341-356.

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

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-75.

MacKay, D. G. (1987). *The organization of perception and action*. New York, NY: Springer Verlag.

Page, M. (2000). Connectionist modeling in psychology: A localist manifesto. *Behavioral & Brain Sciences, 23*(4), 443-512.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56-115.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*(5), 377-500.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*(3), 460-499.

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

Roelofs, A. (1997). The WEAVER model of word-form encoding in spoken production. *Cognition, 64*, 249-284.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review, 110(1)*, 88-125.

Ruml, W., & Caramazza, A. (2000). An evaluation of a computational model of lexical access: Comment on Dell et al. (1997). *Psychological Review, 107*(3), 609-634.

Ruml, W., Caramazza, A., Shelton, J. R., & Chialant, D. (2000). Testing assumptions in computational theories of aphasia. *Journal of Memory & Language. Special Issue: Disorders of language and memory: Implications for cognitive theory, 43*(2), 217-248.

Santiago, J., MacKay, D. G., Palma, A., & Rho, C. (2000). Sequential activation processes in producing words and syllables: Evidence from picture naming. *Language & Cognitive Processes, 15*(1), 1-44.

Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 896-918.

Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology, 41*(2), 101-175.

Wheeldon, L. (2000). Aspects of language production. Hove, UK: Psychology Press.

Wheeldon, L. R., & Monsell, S. (1994). Inhibition of spoken word production by priming a semantic competitor. *Journal of Memory and Language, 33*, 332-356.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception & Performance, 24*(4), 1131-1161.

Note

The author thanks Johannes Ziegler and Stéphane Dufau for their comments on the manuscript.