Errare humanum est.
Refusing to 'appreciate' this fact could be a big mistake!

Michael Zock
CNRS-LIF, Marseille, France
michael.zock@lif.univ-mrs.fr

Abstract

Being able to communicate is vital in our modern world, but this has a price. It requires learning, that is, readiness to invest effort and time. Also, learners must be curious, open-minded and adventurous, i.e. willing to step out of their comfort zone to test unexplored hypotheses. Languages, or more precisely, the process underlying their production or interpretation, cannot be taught. Languages can only be learned, but not in a single trial. In sum, learning a language, i.e. be able to produce it in real-time is a long, error prone process. Note, even native speakers do make mistakes, which generally are errors, as they are based on lack of attention (performance) rather than on lack of knowledge (competence).

I will try to show in this paper why this is so, and how we can use it to our advantage, be it to get a better understanding of the process (language production), or to build tools to support authors, helping them to avoid errors or to overcome them. In sum, I believe that errors, rather than being simply a kind of nuisance that should be avoided (or eradicated) by all means, they have certain qualities we should take advantage of.

Index Terms: speech errors, language acquisition, language production, assistive tools

1. The nature of the problem

Spontaneous speech is a cyclic process involving a loosely ordered set of tasks: conceptual preparation, formulation, articulation (Levelt, 1989). Given a goal one has to decide what to say (conceptualization) and how to say it (formulation), making sure that the chosen elements, words, can be integrated into a coherent whole (sentence frame) and do conform to the grammar rules of the language (syntax, morphology). During vocal delivery (articulation), in itself already a quite demanding task, the speaker may decide to initiate the next cycle, namely starting to plan the subsequent idealational fragment.

In sum, speaking or acquiring this skill is a daunting task requiring the planning and execution of a number of subtasks. Given some goal a speaker must plan what to say (conceptual level) and how to say it (linguistic level), that is

- find the right words (lexicalisation);
- determine an appropriate sentence frame (syntax);
- insert the chosen words in the right place (syntax);
- add function words (syntax);
- perform morphological adjustments (morphology);
- articulate (phono-acoustic level).

While doing all this, in particular, while getting closer to the final form, the must speaker must start to plan the next stretch of discourse (conceptual fragment). Obviously all these stages must not only be well defined and well organized, they must also be properly executed, the output of one step being (generally) the input of the next lower level. Yet, there is more to this. Apart from the number of steps we need to consider the following facts and constraints: (a) discourse is very fast, (b) words must be located in a huge repertoire (mental lexicon); (c) the processor's short-term memory capacities, i.e. the human mind, are very limited (Miller, 1956). We can keep only a limited amount information for a very limited amount of time in the buffer. Yet, many operations are performed precisely on the basis of the information available in this (working) space.

Consider the following problems, both related to space and time. Speaking is basically a sequential process, component b relying on the results of component a. Hence, any hesitation in one component, say, lexical choice, may yield a delay of the next lower component (syntax or morphology). Also, the results of a higher component may need to be revised in the light of results coming from a lower component (retroaction). Correlated to the time problem (delay) there is also a space problem. Any symbol waiting for translation (say, the mapping of a concept into a word) needs to be stored, taxing short-term memory, a very scarce resource.

Last, but not least, speakers may face an expressibility problem, there is a gap (Meteer, 1990) between language and thought: given some message, there are no corresponding linguistic resources (words, syntactic patterns) fitting the planned conceptual structure. In such cases, fall-back strategies must be considered, paraphrases being one of them.

Taking all this together it may come as no surprise, that the route leading from some input (goal) to the final form (output) can be long and full of obstacles (hesitations, errors,...). One may even wonder how people manage to do all this in such a short amount of time given all these constraints. Actually, we do know the answer, at least partially: to make all this feasible, we do not carry out all steps in strict serial order, one process starting only when the preceding one has been fully carried out, rather the process is incremental and semi-parallel (Kempen & Hoenkamp, 1987).

2. Errors in speech

We tend to initiate speech, i.e vocal output, before having planned in all details everything we are going to say. Metaphorically speaking, speakers are players, taking chances, planning opportu-

---

1 For more details concerning language generation architectures, see (deSmedt, Horacek and Zock, 1996).
nistically. Language producers are juggling constraints (Flower & Hays, 1980).

All this shows that speaking is a risky activity. The process is anything but safe, algorithmic or foolproof. Operating often on only partially available information, speaking is potentially very error prone. I will illustrate here why errors may occur at the syntactic or lexical level, and with regard to the latter problem, I will quickly show how we can help authors to recover from them. Finally, I will present a psychologist's work (Dell, 1986) simulating in a connectionist framework how errors can occur when trying to find a word.¹ Yet, before doing so, let me briefly summarize some work on speech error data.

2. The golden route for profilers, model builders and process trackers

With the advent of neurosciences, there are nowadays many ways to 'look' into the human brain.² However, one of oldest techniques is speech-error tracking. It goes back to the end of the 19th century (Meringer & Mayer, 1895; Freud, 1901), and it has often been considered as the golden route to understand the hidden structure of the brain, its organization and functioning allowing it to produce language in real time.

Errors are like goals in a soccergame. They are easily recognized, and so can be their causes, provided they are the last element of the chain doing apparently something 'wrong' (goal keeper, defendant). Errors, or sub-optimal choices are much harder to spot when they are caused earlier on in the process (the midfield, to maintain the analogy with soccer). Yet, this is often the case. As speech error literature clearly shows, errors can be made at any level, that is, by any of the components contributing to the final result (conceptual level, lexical level, syntactic structure, ...).

Speech error data have been used for shedding light on many aspects of language production,³ but they have been used in particular to build a psychologically reasonable model.⁴

2.1. Use errors to build a model

One of the pioneers in this respect is Garrett (1975, 1980) who used errors to build a model, accounting for the process linking an input (message) to its corresponding form (sound). Having studied a corpus of several thousand naturally occurring errors, he noticed that many of them affected content words of similar semantic or syntactic categories (substitution errors: left instead of right; you can cut rain in the trees). Others, though also being content-words, affected only direct neighbors, possibly causing sound exchanges (shinking sips).

It is on the basis of these kind of errors that Garrett built his model, concluding that if two elements of a sentence are both involved in an error, they must be simultaneously available at the stage of processing at which the error occurs. Likewise, if two types of elements are never involved in a given kind of error, then they must be processed at different stages.

According to Garrett, the transition from a message to its corresponding surface form is accomplished via a number of distinct, serially ordered steps, each level involving different kinds of representation. The conceptualizer determines the message to convey. The formulator takes care of processing of lexical, syntactic and morphological information. The articulator transforms the phonetic representation into a series of muscle movements.

Garrett makes very precise claims concerning the stage at which words are generated. At the functional level, the speaker has access to the semantic representation of a content word, but not to its morphological or phonological representation. It is only at the positional level that phonologically specified word forms are accessed and placed in a syntactic sentence frame, to which function words are then added. Note that Garrett was able to confirm certain predictions concerning speech errors. For example, if content words and function words are accessed at a different moments, than they should produce different kinds of errors. This could be shown. Indeed, certain kinds of errors never occur, while others can be elicited experimentally.

Note also that these kind of proposals and concerns are quite different from those of computational linguists, i.e. engineers (Reiter & Dale, 2000). The differences lie in their respective goals, scopes (sentence vs. text) and processing assumptions. Obviously, humans and computers are subject to different constraints (attention span, memory size, and processing speed). For more details on this, see (Bateman & Zock, 2015; de Smoedt, Horacek, Zock, 1996; Dijkstra & de Smoedt, 1996).

Next to building speech production models, errors have been used extensively to get a better understanding of the mental lexicon. One of the questions being asked was, why do speakers have problems to access a word? In order to answer this last question let us take a look at the time course of lexicalization. According to psychologists (Atchinson, 2003; Levelt, 1992; Bock & Levelt, 1994) the human lexical access system has a highly modular organization. Processing is said to take place at two different levels, the level of lemma, which encode the syntactic and semantic properties of lexical items, and the level of lexemes, which deals with their morphological and phonological forms. Put differently, lexical access takes place in two, temporally distinct stages. In the first stage the speaker checks the semantic lexicon for a lemma expressing

¹ Note that in connectionist frameworks words are not located, the parts contributing to it (meaning, form, sound) are activated. Put differently, word forms are not stored as they would be in a computer program.
² For example, recent neuroimaging and electrophysiological studies have shed light on the neural correlates of speaking. These techniques include PET (positron emission tomography), FMRI (functional magnetic resonance imaging), MEG (magnetoeencephalography), and LRP (lateralized readiness potential) analyses. For details, see Roelofs (2004). Another popular method is eyetracking.
the conceptual input.\(^1\) If he can find one, he will take it and consult then the phonological lexicon in order to find then the appropriate phonological form.\(^2\)

Errors can occur at both ends. Yet, since the two stages are independent, errors belong to either category, but hardly ever to both. Errors at the semantic level will yield a wrong lemma (e.g. *hate* instead of *love*), while errors at the phonological level may result in the wrong phonological form. For example, the intended “relegate” may surface as “rengate” or “delegate” (/r/ and /n/ and /r/ and /d/ being phonologically close), “turn on the heater switch” may result in “turn on the sweeter hitch” (Clark & Clark, 1977), etc. As one can see, these words are all phonologically reasonably close to the target word, yet, it is precisely this proximity that prevents the speaker to access the “right” form.

Work on memory has shown that access depends crucially on the way information is organized (Baddeley, 1982). From the speech error literature (Fromkin 1973) we learn, that ease of access depends not only on meaning relations (syntagmatic and paradigmatic associations), but also on linguistic form. Researchers collecting speech errors have offered countless examples of phonological errors in which segments (phonemes, syllables or words) are added, deleted, anticipated or exchanged. Reversals like /animal/ instead of *animal*, or /c teaspoon/ instead of *teaspoon* are not random at all, they are highly systematic and can be explained. Examples such as grastly, a blend of *grizzly* + *ghostly*, or *Fatz* and Kodor (instead of Katz and Fodor) early show that knowing the meaning of a word does not guarantee its access.

I will show in the next subsection what other framework has been used to account for the process, accomodating the data we have concerning language in action (performance).

2.2. Connectionism, a framework allowing to accomodate data of language performance

Language processing can be modeled in many ways (symbolic, i.e. rule-based; hybrid, i.e. statistical processing; connectionist, i.e. neural networks). Of all these, neural networks are particularly interesting for us with respect to modeling certain features of the human mind (process). Rather than taking the computer as a metaphor for the human mind, connectionists take the brain as a metaphor for the computer. For details, see (Bechtel & Abrahamsen, 1994; Pinker, 1984; Spitzer, 1999).

Unlike computers which processes information via a single, central processing unit, the biological brain relies on a great number of interacting elements (neurons). Although each neuron has only limited computing power, taken together neurons are highly noise-resistant. Containing a lot of redundant elements they can cope some extent with damage due to a loss of elements. Hence, rather than bringing the system to a complete stop, this kind of architecture allow for recovery and graceful degradation. This tolerance of gradual or partial breakdown fits nicely psycholinguistic data concerning aphasia, amnesia, speech impairments, speech errors, and the “tip of the tongue” problem (Brown & Mc Neill, 1966; Béroule & Zock, 2009).

Also, connectionist architectures are highly interactive. The multi-directional spreading of activation explains how top-down constraints may bias bottom-up processes. For empirical support, see (Harley, 1982, 1984; Dell, 1986). Taken together one may say that connectionist architectures represent an important effort to overcome some of the limitations of traditional representation of information (linguistic knowledge) and its processing. Apart from simulating learning behaviour, connectionist architectures show greater robustness in dealing with incomplete input and incomplete knowledge sources. Hence neural networks are appealing not only on intuitive grounds, but also in terms of matching empirical data concerning natural language generation: parallelism, competition, etc. are design features often pleaded for by psycholinguists (Stemberger, 1985; Garman, 1990).

In the next two sections I will show what factors may lead to speech errors, beginning with syntax (section 3) to proceed then to word access (section 4). Finally I will show (Dell,1986) how a connectionist model can account for speech errors when trying to access a word.

3. Possible errors when generating syntactic structure

The main function of language is communication. To this end it offers a set of devices allowing the expression of meanings and the relations between them. Lemmata (e.g. *go* vs. *run*), word forms, i.e. lexemes (e.g. *goes* vs. *went*) and syntactic structures (nominalization, relative clauses, ...) being the main means.

Suppose you wanted to express the following idea (a) “beer comes from many different countries”. Having started to produce *beer*, the following afterthought may come to your mind : (b) ‘students drink beer’. These two ideas can be expressed independendy (a-b) or as an embedded clause, ‘b’ occurring inside of ‘a’, yielding "Beer, students buy, comes from very many different places".

While this is not only perfectly correct it is also quite understandable, but things can change quickly. Imagine that a similar situation occurs at the very moment of producing the word ‘students’. Again, many languages allow us to integrate all this information into a single sentence, yielding something like : [Beer,1, students2, [policemen, follow,1], buy2], comes, from very many different places. While being grammatically correct there are various problems, some only potential, others more real, problems that will become more serious as the number of embeddings grows.

First of all, embeddings affects the listener. Indeed, this kind of sentence is hard to understand (in particular if the links between the subject and the predicate are not predictable), as the listener has to store a list of subjects (beer, students, policemen), which she links then in inverse order (LIFO) to their corresponding predicates (follow, buy, come).

This being said, this kind of structure, called center-embedding, may also end up becoming a problem for the speaker, as he must not only make sure to produce the respective predicates in the correct order (beer-come-many places; students-buy-beer; policemen-follow-students), but also produce then the appropriate agreements. Yet, in this particular setting (beer\(_{sing}\) students\(_{plural}\) policemen\(_{plural}\))
chances are that the speaker makes an agreement error, to produce the plural form 'come' instead of the singular 'comes'. This is not a mistake, the speaker knows the rule. It is an error having various possible causes: lack of attention, i.e. divided attention, not tightly controlled allocation of resources (book-keeping); short-term memory constraints (overload) and priming: all but the first noun in the chain are in plural, making the speaker forget to pay attention, i.e. to remember the number of the subject of the main clause ("beer comes from many different countries."). Recursion, its various types and effects on processing have been widely studied in the literature (Yngve, 1960; Dowty et al. 1985). For a good overview see (Christiansen & Macdonald, 2009).

4. Possible errors when trying to retrieve the form of a word (lemma)

In the discussion concerning possible causes accounting for word finding problems I mentioned the fact that words are stored in two modes, by meaning and by form (sound). It is often this latter which inhibits finding the right token: having recomibined inadvertently the components of a given word (syllable scrambling), one may end up producing a word, which either does not exist, or is simply different from the one in mind. This kind of recombination, resulting from bookkeeping problems due to time pressure, parallel processing and information overload, may disturb or prevent the access of words. Hence the usefulness of a tool that allows to revert the process. In order to allow for doing so, it is necessary to represent words not only in terms of their meaning, but also in terms of their written and spoken form.

The fact that words are indexed both by meaning and by sound can now be used to our advantage (Zock & Fournier, 2001). Suppose you presented the system an unknown word (aclibérer), or a word that does not exactly express the intended meaning. The fact that words are coded phonetically allows the recombination of their segments (syllables), hence the presentation of new candidates, among which the user should find the one s/he is looking for. The fact that words are coded semantically allows keeping the number of candidates to be presented small. The list of potential candidates will be filtered according to semantic criteria (domain). Hence the phoneme /vin/ would yield vin, vaire, vingt or vint depending on whether the domain were "food (wine), battles (to win), digits (twenty), or movement (to come)."

As one can see, if phonological permutations may inhibit the access of a given word, they may also enhance its access: used deliberately and in a controlled way, they may allow the system to override illegal combinations: presenting the user with other alternatives, among which s/he may find the one s/he was looking for. Strange as it may seem, this kind of technique has been used for quite some time in the area of spell checking.

The component described here below has two basic mechanisms for correcting spelling errors: anacodes and phoneticodes (for details see Fournier & Letelier, 1990). The former computes an access key for finding the right word. Since an anacode is equivalent to the set of letters composing the word (for example, the anacodes of "calibérer" and "aclibérer" are identical), erroneous order of letters is a non-issue. The system would still find the right word, provided that there is such a candidate in the dictionary, and provided that the user did not omit, add or replace some character(s) with other characters. For example, if the input were aclibérer instead of calibérer, the system would have no difficulty to find the target word (calibérer), since both words are composed of the same set of letters. If the user added letters outside of the anacode, the system would need several runs to check alternative spellings by making local variations (delete or add a character by making systematic permutations). Unlike most systems, this technique allows to deal with spelling errors occurring at the beginning of a word.

The second technique (phoneticodes) consists in converting graphemes into phonemes, which allows the system to deal with spelling errors due to homophony (two different words having the same spelling), a very frequent phenomenon in French. For example, the system would be able to deal with errors like hypoténuse instead of hypoténuse. Put differently, if the user presented the word hypoténuse, while he was trying to say hypoténuse, the system, rather than remaining silent, would present the word he was looking for, namely hypoténuse. If the system cannot find directly a candidate, it will perform local changes by performing permutations of phonemes or syllables. Hence it would have no problem to find the word "poutau" (pole) instead of "topo" (topic), both words being composed of the same syllables (/po-to/), the only difference being their order.

The situation is more complex and may even become intractable if extraneous material is added, or if the correction yields an existing word, yet different in terms of meaning from the one intended. Suppose that the target word were "maison" (house), while the user typed /masson/. Relying on the phoneticode, the system might suggest "maçon" (bricklayer), a word that exists, but which is not at all what was intended.

5. Error simulation via a connectionist network

In this last section I will briefly present a connectionist network simulating speech error when retrieving words. For connectionist approaches dealing with syntactic problems, see (Kalita & Shastri, 1987; Ward, 1994). The dominant psycholinguistic theories of word production are all activation-based, multilayered network models. Most of them are implemented, and their focus lies on modelling human performance: speech errors or the time course (latencies) as observed during the access of the mental lexicon. The two best-known models are those of Dell (Dell, 1986) and Levelt (Levelt et al. 1999), which take opposing views concerning conceptual input (conceptual primitives vs. holistic lexicalized concepts) and activation flow (one-directional vs. bi-directional).

---

1 The best known work illustrating this problem is the Stroop effect, named after its author (Stroop, 1935). The classical experience is a demonstration of interference in the reaction time of a color-naming task. When the name of a color (e.g., "blue", "green", or "red") is printed in a color not denoted by the name (e.g., the word "red" being printed in blue ink rather than red ink), naming the color of the word takes longer and is more error prone than when the two, the color of the ink and the color name, match.
Dell’s model (fig. 1) is interactive-activation-based. Starting from a set of conceptual features, the system generates a string of phonemes. Information flow is bi-directional, that is, lower level units can feed back to higher-level components, possibly causing errors similar to the ones people tend to make. For example, the system might produce rat instead of the intended cat. Indeed, both words share conceptual and phonological components. Hence, both of them are prone to be activated at the same time, leading thus to potential conflicts. At the conceptual level (from the top) they share the feature animal, while at the phonological level (from the bottom) they share two phonemes. When the lemma node cat is activated, any of its phonological components (/k/, /æ/, and /t/) is activated too. The latter two may feed back, leading to rat, which may already be primed and be above baseline due to some information coming from a higher-level component. Note that this model can produce various other kinds of speech errors like preservations (e.g., beef needle soup), anticipations (e.g., cuff of coffee), etc.

Given the distribution of word errors, Dell argues that speech generation relies both on retrieval (phrases, phonological features, etc) and synthesis (word/phoneme and possibly morpheme combinations). It is precisely during this latter process that the mechanism may make mistakes, swapping, anticipating or reusing elements.

### 6. Conclusions

Errors are a necessary evil of learning and progress. Being unavoidable for learning and processing of language in real-time, they are also a gateway or window to the mind. Hence they are precious information for teachers, researchers and language practitioners (speakers) alike. Put differently, rather than being only a form of nuisance, errors can be considered as a valuable (and exploitable) resource. They allow us not only to shed light on the mental processes, but also identify the needs and cognitive states of the language producer, which is precious information for those who want to explain the process, help its acquisition (trachers) or wish to build the needed tools (engineers).

Note that in rule-based systems errors are due to ignorance. In connectionist architectures they emerge as a side effect of weak connections. Hence the question, how to boost weights? One answer to this resides in the increase of connections (associations) which allows not only to recover from errors, but also to overcome the well-known tip-of-the-tongue problem ((Brown & Mc Neill, 1966)), that is, the blockage a speaker may experience when she looks for a word she knows, but is not able to activate.

While errors should be avoided to begin with, there is no fool-proof way for doing so. In addition this strategy may turn out to be counterproductive (stress, inducing silence, minimal risk taking, ...). Since there is no way to eradicate them, it seems wiser to accept them and to learn how to recover from them if ever they do occur. One of my goals in this paper has been to show how this can be achieved. While I have focused here only on word access based on form, one could also think of word access based on meaning or based on associated terms. This is something I am currently working on (Zock et al. 2010; Zock & Cristea, 2014).

### 7. References


