Learn to *speak* and to *write*, learn to *use* your *mind*
The relevance of *automatic text generation research* for people

Michael Zock
LIMSI-CNRS, B.P.133,
91403 Orsay, France
zock@limsi.fr

Abstract

The aim of this talk is to show to what extent the work on ***t**ext **g**eneration **b**y **c**omputer* (TGBC) does not address some of the fundamental problems ***people*** struggle with when generating language (TGBP). We will substantiate this claim by taking two tasks on which a lot of research has been carried out during the last 15 years: *discourse planning* and *lexicalisation*.

1 Discourse planning

While a tremendous amount of work has been done on the generation of *coherent discourse*, little if any has been devoted to *writing*. As a result, many fundamental problems have been overlooked or have been dealt with on the basis of wrong assumptions. Also, little, if any of the results achieved in the TGBC framework can be reused in the classroom or in the context of an intelligent writing-aid (tools for assisting the writer to structure her/his thoughts: *outline planning*). Let us consider some of the reasons why this is so.

• ***Top-down processing:*** in the TGBC-community texts are generally processed top to bottom. Given some goal one looks for *data* (messages) and *structures* which integrate them. While this is a clever way to handle the problem, it does not give a precise reflection of the writers' situation. First of all, it is not true that *content* and *structure* are always determined simultaneously, an assumption accepted since Moore & Paris (1993). Secondly, writers generally switch between *data-driven* (brainstorming) and *structure-driven* processing (outlining). Thirdly, there is a triangular relationship between *messages*, *structures* and *goals* (or *effects)*, changing any of them can affect the others. Yet, at present we do not have the faintest idea what *effect(s)* a specific propositional or conceptual configuration (order of messages) might produce.

• ***Lack of a Conceptual Structure Theory* (CST)**: messages tend come to our mind in any order and **without** *exhibiting* their *potential* links. We have to ***discover*** these later, and to reorganize the former in order to *reveal* the structure to the reader. *Writing is thinking*. These last three points are crucial, yet none of the existing theories (schema, RST) is really able to take them into account. Just imagine how complex it is to recognize the fact that there is a ***causal*** link between two events. We don't have a solid theory of *causality*, leave alone a method of operationalizing it (i.e. infer this kind of link solely on the basis of the intrinsic features of the events involved).

• ***Interaction***: As we all know, texts have structure. This latter is generally the result of *discourse planning* (schemata or RST-based) or *reasoning* (chain of inferencing), in which case the structure emerges as a side effect. The major shortcoming of all these techniques is that they do not model the *interaction* between the *conceptual data* (ideas, messages), the *text structure* and the *rhetorical effects*: (all) the data to be communicated and the global discourse

goal are generally given with the input.[1] The problem of reconciling mismatches between *data* and *structure*,[2] and the problem of variable *rhetorical effects/goals* as a function of various linearization strategies is not addressed at all.[3]

2 Lexicalisation

Lexicalisation amounts mainly to *searching* and *choosing:* one has to *find* lemmata, matching a given conceptual chunk, and then one has to *choose* among them. While much emphasis has been given to the notion of *choice*, far less attention has been paid to the *search* mechanisms (or access strategies). I will present during my talk some preliminary results concerning a system that is meant to help people to overcome the *tip-of-the tongue problem*, a well known stumbling block in real-time processing: we know *what* we want to say, we know that we *do* know the word, yet we cannot *access* it (Brown and Mc Neill, 1966).

If the fundamental role of a dictionary in NLG is obvious, it is less evident as to the principles governing its compilation. A good dictionary is a place with a *lot of information, structured* in such a way that the relevant information is easily ***accessible*** when needed. In other words, what counts is *'what* is *in* the dictionary' (*content*) and '*how* the information is *organized* (meaning, form, sound). These two factors are not sufficient though: *access* depends not only on the *structure* of the lexicon (organisation), but also on the *efficiency* of *search strategies*, an issue not addressed at all by the generation community. As a matter of fact, from a strict *computational linguistic point of view*, the whole matter may be a non-issue. However, the problem does become relevant when we look at generation as a *machine-mediated process* (people using a word processor for writing) or from a *psycholinguistic* point of view: word access in writing or spontaneous discourse.

• **The speaker's problem :** *choosing* **words,** *finding* **them or** *both* ? Obviously, there is more to lexicalisation than just *choosing* words: one has to *find* them to begin with. No matter how rich a lexical database may be, it is of little use if one cannot access the relevant information in time. *Access* is probably THE major problem that we have to cope with when trying to produce language in real-time (in spoken or written form). As I will show during my talk, this is precisely a point where computers can be of considerable help.

Work on *memory* has shown that *access* depends crucially on the way information is organized, yet the latter can vary to a great extent. From *speech error literature* we learn, that ease of access depends not only on *meaning relations*,—, i.e. the way words are *organized* in our mind),— but also on *linguistic form* (letters, phonemes). Researchers collecting speech errors have offered countless examples of *phonological errors* in which segments (phonemes, syllables or words) are added, deleted, anticipated or exchanged (Fromkin, 1993). The data clearly show that knowing the *meaning* of words does not guarantee their *access*.

The work on speech errors also reveals that words are *stored* in at least two modes, by *meaning* and by *form (written, spoken)*, and it is often this latter which inhibits finding the right token: having inadvertently recombined 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

---

[1] While in Moore & Paris (1993), the messages are not given, the goal is : it cannot emerge as a side effect.

[2] What shall we do if not all the data can be integrated, or if we lack data for filling all the slots of a chosen structure? Shall we keep the structure and look for more data, or use a different structure as it integrates more of the data?

[3] One of the reasons for this is that we do not have a clear understanding concerning the mapping between different *conceptual configurations* and their corresponding *rhetorical effect*(s). If we did, we could use them bidirectionally (for analysis and generation).

access of the right word. Hence the usefulness of a tool which allows the process to be reversed. In order to allow this to be done, 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* could now be used to our advantage. The *phonetic* coding of words 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.[4] The fact that words are coded *semantically* keeps the number of candidates to be presented small.

Conclusion

I have tried to illustrate briefly to what extent we have neglected the human factor in our work. I have also attempted to show how a simple computational method (combinatorics and filtering) can be used to bridge (one of) the gap(s) between TGBC and TGBP: text generation by people.

References

Roger Brown and David Mc Neill. 1966. The *tip of the tongue* phenomenon. Journal of Verbal Learning and Verbal Behavior , 5, 325-337

Viktoria Fromkin. 1993. Speech Production. In *Psycholinguistics* edited by Jean Berko-Gleason & Nan Bernstein Ratner. Fort Worth, TX: Harcourt, Brace, Jovanovich

Johanna Moore and Cecile Paris. 1993. *Planning text for advisory dialogues: capturing intentional and rhetorical information*. Computational Linguistics, 19(4).

[4] The assumption is that speakers produce words that *formwise* are reasonably close to the target word. A fact that is supported by psycholinguistic evidence.