

The Power of Words in Message Planning

Michael Zock

Language & Cognition
LIMSI - C.N.R.S., B.P. 133
91403 Orsay, France
zock@limsi.fr

Abstract: Before engaging in a conversation, a message must be planned. While there are many ways to perform this task, I believe that people do this in the following way, in particular if the message is going to be long: first an outline is planned (*global, or skeleton planning*), which is then filled in with details (*local planning, elaboration*). Planning proceeds thus from general to specific (breadth first), that is, sentences are planned incrementally by gradual refinement of some abstract thought rather than in one go (one-shot process) where every element is planned down to its last details.

While *global planning* is largely language independent, *local planning* can be language dependent: the dictionary acts a *mediator*, interfacing language and thought. Given the fact that words can be used to specify non linguistic thought, there is feedback from the *lexical* to the *conceptual* component. This being so, dictionaries may play a fundamental role in guiding and potentially modifying non linguistic thought. If my view is correct, this could have implications on the design of generation architectures: instead of separating message planning and realization, viewing the process as being strictly sequential, we could allow for feedback loops (interleaved process), whereby the linguistic component could feed back to the conceptual component.

1 Introduction

A major step in natural language-generation (NLG) consists in choosing content words for expressing the planned message. While this sounds self evident, it contains at least two assumptions that are easily overlooked: (a) thought precedes language; (b) thought is entirely encoded or specified before lexicalization takes place. My contribution in this paper consists in providing evidence for the following three claims: (a) thought is underspecified at the onset of lexicalization; (b) language can feed back on thought, i.e. words can specify the conceptual component; (c) our mental dictionaries are the interface between language and thought.

NLG has often been viewed as a two step process. During the first (deep generation) conceptual choices are made (content determination, discourse planning), during the second (surface generation) linguistic operations are performed (word choice, determination of syntactic structure). While this kind of decomposition has proven useful for practical purposes, —dividing

the process into separate components increased the control, — it has also encouraged researchers to build into their systems wrong assumptions: content is generally determined in one go (one-shot process), and information flow is one-directional, going downwards from the conceptual level to the linguistic level. As I will show by taking an example from the lexicon, both these conclusions are ill-founded, as they suggest that there is no feedback between the different components.

2 A naive view of generation

NLG can be viewed as the process of finding a linguistic form for a given conceptual fragment. Obviously, there are certain dangers with this view. First of all, the *order of thought*, i.e. the order in which conceptual chunks become available, and the *utterance order*, i.e. the order in which words have to be uttered in a given language is not necessarily the same. Second, words cannot be *directly* mapped on their conceptual counterpart, that is, there is no one-to-one correspondance between concepts and words: a given word may express more than a single concept: *leave vs. go away; unhappy vs. not happy* (for more details see Nogier & Zock, 1992). Third, there is no feedback from the *linguistic* to the *conceptual* component. In the remainder of this paper I shall concentrate precisely on this latter problem by showing the interaction between these two components.

3 How is content planned?

Suppose we wanted to produce the following sentence: «When the old man saw the little boy drowning in the river, he went to his canoe in order to rescue him». Figure 1 here below can be considered as the underlying planning tree. But how is such a tree built?

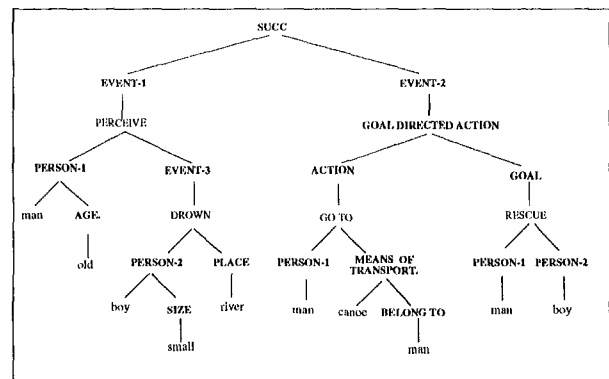


Figure 1

There are several good reasons, —psychological and linguistic,— to believe that this sentence has not been planned from left to right and in one go. Psychological reasons: The sentence is simply too long for a speaker to hold in short-term-memory all the information to be conveyed. It is highly unlikely that the speaker has all this information available at the onset of verbalization. The need of planning, that is, the need to look ahead and to plan in general terms, increases with sentence length and with the number and type of embeddings (for example, center embedded sentences). There is also good evidence in the speech error literature for the claim that people plan in abstract terms. False starts or repairs, like « I've turned on the stove switch, I mean the heater switch » suggest that the *temperature increasing device* has been present in the speakers mind, yet at an abstract level (see Levelt, 1989; Fromkin 1993). Linguistic reasons: as mentioned already, the *order of words* does not necessarily parallel the *order of thought*. For example, the generation of the first word of the sentence here above, —the temporal adverbial "when",— requires knowledge of the fact that there is another event taking place. Yet, this information appears fairly late in the sentence.

Figures 2 and 3 here below illustrate a reasonable way to plan such a message. Starting with something fairly general like, —there are two temporally related events: EVENT-1 preceding EVENT-2 (step-1),— the speaker expands gradually each element (steps 2-8).¹

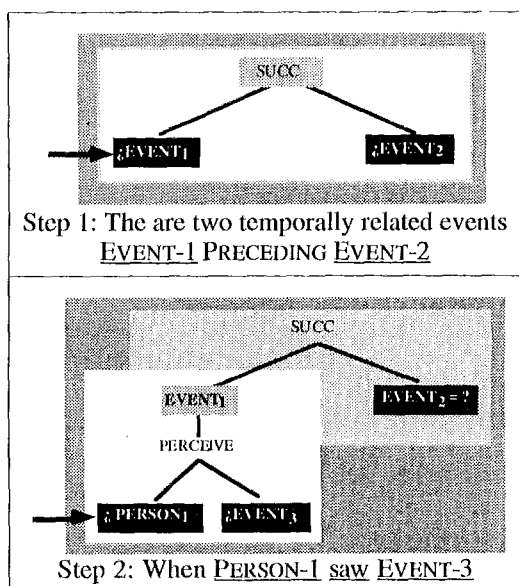


Figure 2

¹ I will use the following conventions in the figures: *text with gray background in white box*: currently expanded element; *white inverted question mark*: not yet fully specified element; *gray box*: processed elements; *pointer*: element to be elaborated.

My comments under the figures should be read in the following way: *underlined element*: currently obtained result; *capital letters*: variable, hence not yet fully specified element.

Step-2: Since there are two events, the speaker has to choose. Let's assume he wants to begin with EVENT-1. This allows the generation of the first element "when", and the fact that there is a PERSON who saw some EVENT. Both *person* and *event* are still unspecified elements, hence written in capital letters. Having two unspecified elements (PERSON, EVENT-3) we have to choose.

Step-3: Suppose we decided to elaborate PERSON. This could yield something like *man* + ATTRIBUTE, meaning, that the person who sees the event is a *man* (a terminal element) whom we want to describe further by providing a characterizing attribute. Please note, that I consider both *man* and the variable ATTRIBUTE as sister nodes, hence, in principle there could be a linearization problem. Whether predicates (in our case, the attribute) can, or are determined before the argument they qualify (here *man*) remains an empirical question. The situation seems clearer during lexicalization where a head noun may constrain an adjective, hence the noun has to be generated first (collocational constraint).

Step-4: During this step we decide on the attribute, the result might be "old". One could object that an intermediate step is necessary in order to decide on the kind of attribute (size, age, etc.). This is correct, but for reasons of economy (size of the figures), we've skipped this step.

Step-5: During this step we elaborate the node EVENT-3 which yields : someone drowns somewhere, where the *person* and the *place* are still unspecified. Again we have to choose which element to elaborate. Suppose we started with PERSON-2.

Step-6: This could yield *boy* + ATTRIBUTE. Boy being a terminal element it needs no further refinement, but we do still need to specify its attribute.

Step-7: If we were to characterize the *boy* in terms of size, we might get "small".

Step-8: The instantiation of the variable (PLACE) might yield *river*.

Having completed the description of EVENT-1 we still have to specify EVENT-2. We will leave this as an exercise for the motivated reader (bon courage!)

4 Possible implications

Let us see some of the advantages of our approach. Top-down, left-to-right expansion gives the speaker a good control over the whole process, minimizing the danger of forgetting some information because of memory overload. Of course, the tree can be built in different ways, top-down, bottom-up, or by combining both methods.² For related approaches on

² I believe, that the way how the tree is built depends on whether the speaker has at the onset of message planning a clear picture of the object or scene to describe, or whether he has to build it from scratch. In the first case he could build the tree from top to bottom.

bidirectional tree growth, see the work on *segment grammar* (de Smedt & Kempen 1991), or *Tree Adjoining Grammar* (Joshi 1987).

If my line of reasoning concerning message planning is correct, —namely that planning is basically a two-step process where first a skeleton is planned (*general plan*), and then its constituents (*specific plan*),— then this should have consequences on the overall architecture of generators, as well as on the information flow (control, process). We shall see this in the next section on lexical choice. The basic question that arises in this context is the following: when do we process what? This suggests the corollary questions: Is all information pertaining to a given module processed in *one go* (one shot), or are objects *gradually refined* (several passes)?

For example,

1. is *everything* related to meaning processed *entirely* and once and for all — hence messages can neither be changed nor be refined — or
2. are the objects to be talked about only specified to the degree to which they need to be at this stage of the process (limited commitment planning)?

Actually, it is quite easy to find arguments in favor of this second hypothesis. Why should we care at an early stage of the process about details, if we aren't sure at that point, whether the global message will meet the goal defined? Put differently, why care at that stage about the specificity of a word or a referent (how great a detail to give in order to characterize an object) if we aren't even sure whether the planned message contains all and only the information we wish to convey? In a similar vein, why bother about style, spelling and punctuation, and so forth, if syntactic structure is likely to be changed? The point I am trying to make is that, people probably start by planning things globally, filling this plan with details at a *later stage* (local planning). We cycle through the same kind of process but at different levels of detail. Having built a global plan we flesh it out with details as soon as this becomes necessary. The speech error literature abounds with examples supporting this point of view. Blends, substitutions, or speech repairs like '*He conquered Babylon, the great Alexander*', seem all to suggest, that the speaker plans his message in abstract terms. For details see (Levelt 1989; Fromkin, 1993). Genuine examples, as the following from Maclay and Osgood (1959:25) clearly show that the speaker has not necessarily everything planned at the onset of articulation: longer stretches of spontaneous discourse are full of pauses, stutters and mistakes:

« As far as I know, no one yet has done the / in a way obvious now and interesting problem of [pause] doing a / in a sense a structural frequency study of the alternative [pause] syntactical [uh] / in a given language, say, like English, the alternative [uh] possible structures, and how / what their hierarchical [pause] probability of occurrence structure is. »

5 Lexical choice as pattern matching

Having planned the underlying content, let's see how the dictionary may feed back on the conceptual component. The process of finding words for conceptual structures can be viewed as lexically mediated, hence indirect structure mapping: conceptual fragments are mapped onto words *via* the lexicon (see figure 4), the latter serving as an interface between thought (concepts) and language (words).

This view raises a number of interesting problems: (a) What shall we do if not all of the planned message can be expressed by the words available at a given moment? Should we backtrack (try to find another word); change the underlying meaning (replan

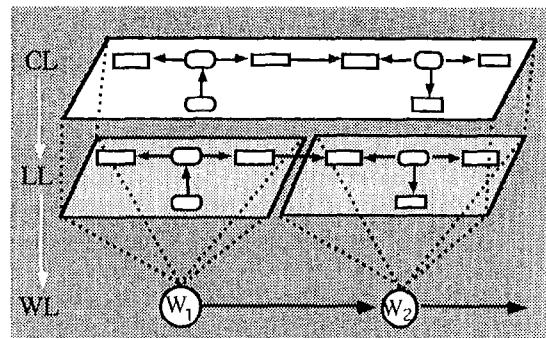


Figure 4: The lexicon (LL) as mediator between the conceptual level (CL) and word level (WL).

the content by adding or deleting some information); or carry over to the next cycle (word or sentence) the specific part that couldn't be expressed (*carry-over phenomenon*)? For an attempt to solve this problem, see Nicolov et al. (1996). (b) On what basis do we choose among a set of potential candidates? (c) What information is available at the onset of lexicalization (all or only part), i.e. at what moment is the word's underlying meaning fully specified?

In order to answer these questions, let me take an example. Suppose we wanted to express the fact that a person moves on a given surface in a given direction (see G1 in Figure 5 next page, or, the left branch of second EVENT of our example in section 3 on message planning: *...he went to his canoe...*).³ Obviously, there is more than one way. According to the input specifications we could consider either '*to move, to swim, to walk*', or '*to run*'. All these words express the notion of movement. Yet, not all of them fit equally well the initial message, and for quite different reasons. While the second candidate ('*to swim*') is simply in contradiction with part of the initial specification (location: ground), the last one ('*to run*') expresses more than the initial message planned. Now, what could motivate the choice between the two remaining candidates, '*to move*' and '*to walk*'? Both of them are subgraphs of the utterance graph, that is, both of them express part of the message planned. Yet '*to walk*' expresses more of

³ Figure 5 is to be read counterclockwise.

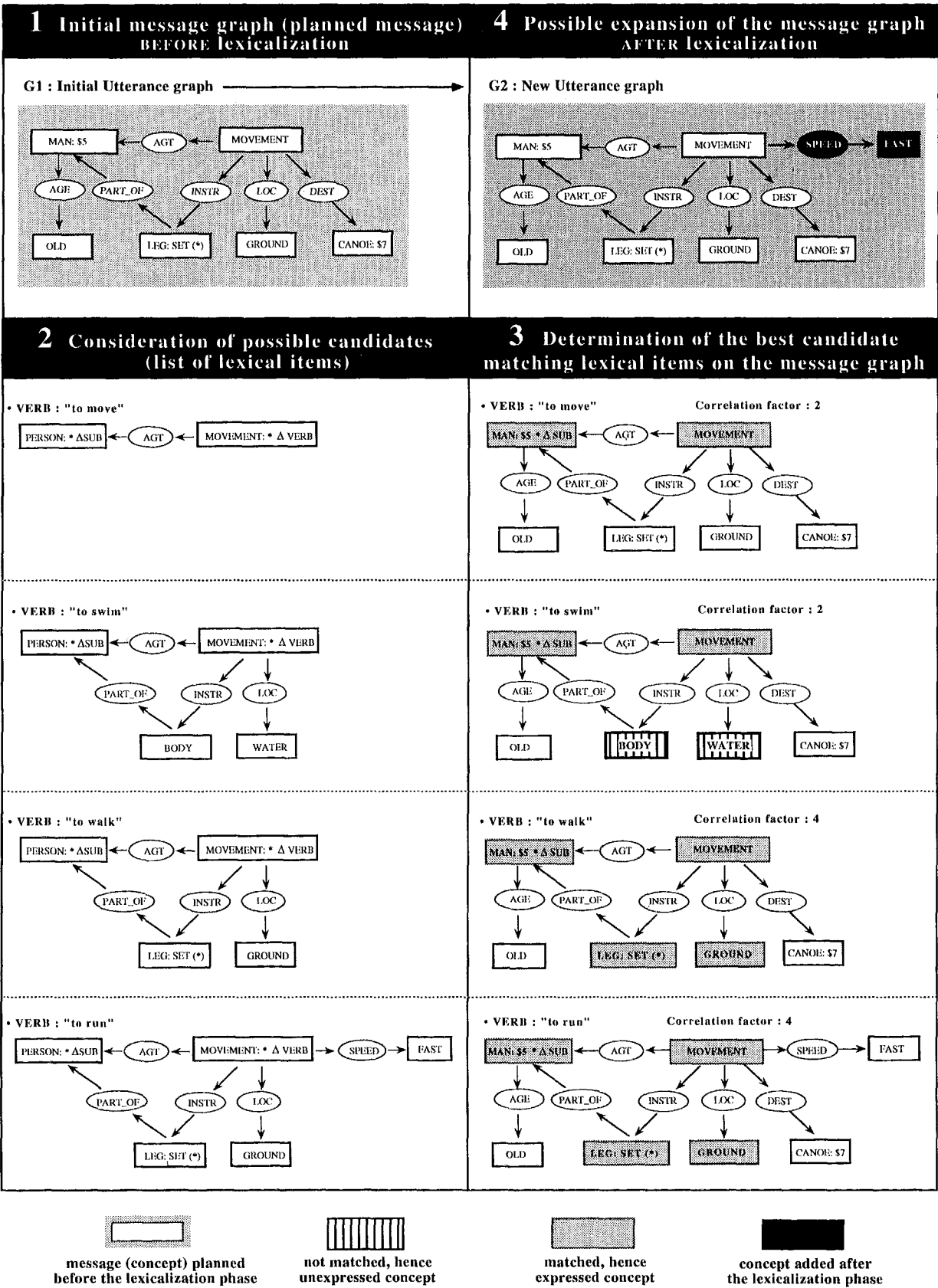


Figure 5

the message graph, hence, all things being equal, it will be preferred.

Actually, the choice of a specific word will depend to a large extent on pragmatic factors like speaker's goals, time and space constraints, hearer's expertise, and so forth. Of course, all these factors can be interrelated. If conciseness is what we are looking for, then the most specific word ('to walk') is to be preferred to the more general term ('to move'). The reason is simply economy: more of the message is expressed by using the given resources, hence time and space is gained. There is a limitation though. The words chosen have to match the user's expertise. A cooperative speaker uses dense words⁴ or technical terms ('computer', 'space shuttle') only for people whose lexical competency allows them to understand their meaning, otherwise he will decompose these words.

Now let me get back to the last candidate ('to run'), as it raises an interesting problem. The word's underlying meaning is not in contradiction with the initial message. Yet it does not express exactly what was planned. Actually, it expresses something more: the notion of speed.⁵ This being so, the question arises, whether 'to run' can be considered as a valid candidate. I believe that it does, provided that the additional information (speed) is consistent with some belief about the state of the world, and that the speaker considers it worthwhile mentioning. If that is so, then we have here evidence for feedback of the lexical component to the conceptual component. The described situation may seem somehow artificial, yet, I believe it occurs quite often in spontaneous discourse, even if we are not aware of it. When we get to the point of choosing a word, it is the power of the language that drives us to say something that was not initially planned. Of course, we are free not to mention the additional piece of information, but this is not really the point. The point is, that not all information ultimately expressed by a word has been available in the speaker's mind at the onset of lexicalization.

6 Discussion

The approach taken here raises an interesting problem. When choosing words we express not only a given meaning, but we may end up adding to the conceptual structure (message) meanings that initially we had not planned. While there are good reasons to believe that at an early stage of processing the underlying meaning of words is underspecified and

⁴ Dense words are generally abstract words like *inflation-rate*, *superstition*, *belief*. They carry a lot of information.

⁵ Please note that this example serves only for illustrative purposes. Actually, 'to run' and 'to move fast on the ground' do not refer exactly to the same kind of locomotion. For more details, see Nogier & Zock (1992).

abstract, hence language independent, the question arises whether the information added during the expansion phase is not *language specific* to a large extent. Much of the information to be added is precisely information required by the lexicon. While it is possible that for syntactic and other reasons we cannot use or retrieve the word planned, it is less obvious why one would want to add at this stage information for which we don't have words (see also Meteer, 1990). Obviously, subscribing to our reasoning is taking a stance with regard to the *language-thought* debate (Whorf, 1956), namely, language may drive thought. Surprisingly, this is all the more likely as we get close to the surface, that is, relatively late in the process.

7 Conclusion

In this paper I have tried to give evidence for three claims, namely that thought is not completely specified at the onset of lexicalization; that there is feedback from the lexical to the conceptual component; and that the dictionary plays a fundamental role in guiding and potentially modifying non-linguistic thought. Hence, it is during the lexicalization phase that language can play this important role over thought. This is what I meant to say, when I wrote in my title: the power of words in message planning.

REFERENCES

- K. de Smedt & G. Kempen: Segment grammar: a formalism for Incremental Sentence Generation, In, Paris, C. & W. Swartout & W. Mann (Eds.). *Natural Language Generation in Artificial Intelligence and Computational Linguistics*, Kluwer Academic Publishers, Boston, 1991
- V. Fromkin: Speech Production. In J. Berko, Gleason & N. Bernstein Ratner (eds.) *Psycholinguistics*. Fort Worth, TX: Harcourt, Brace, Jovanovich. 1993
- A. Joshi: Tree Adjoining Grammars and their Relevance to Generation, in: Kempen (Ed.). *Natural Language Generation: New Results in Artificial Intelligence, Psychology and Linguistics*, Martinus Nijhoff Pbs, Dordrecht, 1987
- W. Levelt: *Speaking : From Intention to Articulation*. Cambridge, Mass.: MIT Press; 1989
- M. Meteer: The 'Generation Gap': the problem of expressibility in text planning., BBN Rep. n° 7347, 1990.
- J.F. Nogier J.F. & M. Zock: Lexical choice by pattern matching. In *Knowledge Based Systems*, 5 (3), 1992.
- N. Nicolov, C. Mellish and G. Ritchie, Approximate Generation from Non-Hierarchical Representations, 8th International Workshop on Natural Language Generation, Herstmonceux Castle, 13-15 June 1996.
- B. Whorf: *Language, thought, and reality*. Cambridge, Mass., MIT Press, 1956