

Converting an electronic dictionary into a drill tutor

Michael Zock and Julien Quint

LIMSI-CNRS, Groupe LIR, B.P. 133, F-91403 Orsay, France
National Institute of Informatics, Hitotsubashi 2-1-2-1312, Chiyoda-ku, Tokyo 101-8430, Japan
zock@limsi.fr, quint@nii.ac.jp

Abstract

Electronic dictionaries have a potential which has been largely underexploited, in particular with respect to foreign language learning. Adding certain functions to them would considerably extend their usefulness. For example, such enhanced dictionaries could support people in their tasks of *reading* (in a script different than your mother tongue), *writing* (reveal the word that is on the tip of their tongue), *memorising* words, or *automating* basic syntactic structures. We will focus here on this latter aspect. That is, we will show here how to convert an electronic dictionary into a drill tutor or exercise generator, or, more precisely, a goal-driven, template-based sentence generator. The examples given are for learning Japanese, yet the approach taken is general enough to allow for accommodation, that is, the computer program could be adapted to other languages, which is precisely our goal.

1. Introduction: becoming fluent, a challenge for any language learner

Text- or discourse-production basically consists in *determining*, *organizing* and *translating* content in order to achieve specific communicative *goals*. We shall be concerned here only with the last component, the translation of a message into its corresponding linguistic form. To do this *naturally*, that is, at a certain pace, without making too many mistakes, is both a challenge and an achievement for any language learner. To this end many people rely on pattern practice. This seems a good strategy, in as patterns allow us to express even quite complex thoughts in one go.

We will show in this paper how an electronic dictionary can be converted into a drill tutor, helping students to achieve proficiency (fluency), while learning (memorizing) at the same time basic syntactic structures. To this end we associate discourse goals with syntactic patterns, which are instantiated then with a set of words, to be chosen by the student.

While fashionable in the 60ies, pattern drills have been criticized ever since then.¹ But as so often, the baby was

¹ Actually, the idea of patterns or schemata is not new. It has a long tradition in *philosophy* (1), in *sociology* (2), in *structural-* and in *text-linguistics* (3-7), in *psychology* (8-14) and in *artificial intelligence* (15-17). While Bartlett, Schank/Abelson, van Dijk and Rumelhart identified patterns on the text- or discourse level (schemata, scripts, macro-structures, story grammars),

thrown out with the bath water. Yet, there is little doubt, pattern drills are quite useful when learning a new language. They promote not only fluency, but they also liberate the speaker, allowing him to concentrate on the *content* rather than its corresponding linguistic *form*.

2. Motivation of this work

There are several reasons motivating our work. We are both involved in a multilingual dictionary project, Papillon.² One of its goals is the building of a huge, multilingual lexical data-base (English-French, several oriental languages) to extract subsequently digital bilingual dictionaries.

Two questions arising in this context are the following:

- (a) how to capitalize on this resource, that is, how to exploit the data-base?
- (b) how to motivate other people to contribute to its creation?

The answer to the first question is that, rather than considering the dictionary as a dedicated component for a single task within a larger system (parser or generator), we consider it as resource to be used for many tasks like reading, writing, memorization of words or automation of syntactic structures.

The answer to the second question is as follows. One way of motivating people to contribute to the PAPILLON data-base, is to offer them a service they might be interested in. In our case, we'd like to show them what can be done with the word they have contributed, by generating sentences in which the newly contributed word occurs. Obviously, in order to allow for this, we have to add to PAPILLON a function absent in conventional, electronic dictionaries, namely, a sentence generator.

Since both authors of this paper are interested in learning Japanese, having at their disposal such a tool fits their personal needs. Yet, this kind of need is felt by many others, hence the idea to build such a tool and to make it available on the web which is our goal.

Harris and Fries dealt with sentence patterns. The idea of linguistic patterns has also been extensively used in the classroom, where pattern-drills have been a major teaching strategy especially during the sixties, when behaviorism was at its peak (18-19). Things have changed radically after Chomsky's devastating critique of Skinner's book *Verbal Behavior* (20-21).

² <http://www.papillon-dictionary.org/Home.po>

3. A goal-driven sentence (or exercise) generator based on templates

As mentioned already, our goal is to build an intention-driven, template-based sentence (or exercise) generator. As we all know, languages have certain regularities which can be captured as patterns.

<p><PERSONAL PRONOUN> wa <NAME> <TITLE> desu ka ? Anata wa <u>Tanaka</u> san desu ka ?</p>
<p><DEM. PRONOUN> wa <OWNER> no <OBJECT> desu ka? Sore wa anata no <u>kuruma</u> desu ka ?</p>

Figure 1 : Template and example in Japanese

Of course, patterns are language dependant. Figure 2 shows how the same information is expressed by quite different patterns according to the language.

<p><PERSONAL PRONOUN> wa <NAME> <TITLE> desu ka ? Anata wa <u>Tanaka</u> san desu ka ?</p>
<p>Be <PERSONAL PRONOUN> <TITLE> <NAME>? Are you Mr. <u>Tanaka</u> ?</p>

Figure 2 : Template and example in English

Yet, patterns may also depend on goals (conceptual, pragmatic). Figure 3 shows some patterns for achieving the goals of *comparison* and *definition*, while the patterns in figure 1 and 2 express information concerning *identity* and *possession*.

<p>Comparison <A> is to as <C> is to <D> <i>Good is to light as evil is to darkness</i></p>
<p>Definition A <X> is a <Y> that <VP> <i>A computer is a machine that processes information</i></p>
<p>Definition / Comparison 1) A <X> is somehow like a <Y> <i>A cat is somehow like a tiger</i> 2) A <X> is a sort of <Y> <i>A bicycle is a sort of a vehicle</i></p>

Figure 3 : Templates triggered by goals

The fact that templates are associated with goals allows the user now to start generation from an intention rather than from a template.³ A goal triggering one or more templates, the user is meant to decide on the words with which he'd like the template to be instantiated with (see figure 4).

³ While not being very natural, it is the norm in most classes based on pattern drills.

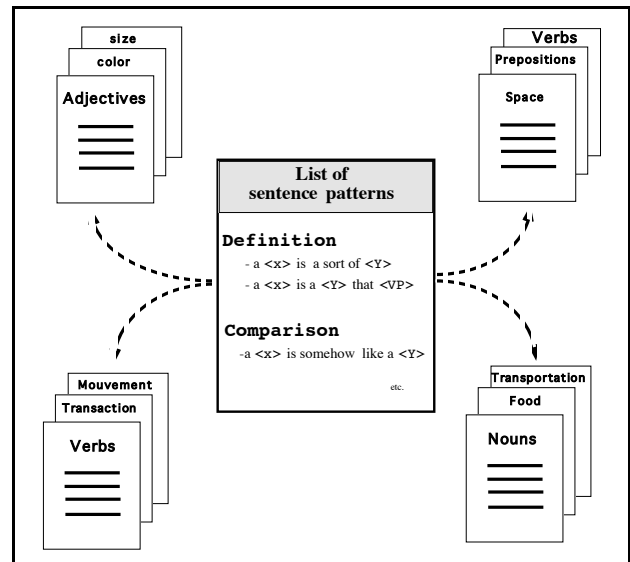


Figure 4 : Sentence generation based on goals and template filling with words

4. An exercise generator tailored to the users' needs

Pattern drills, as most of us have experienced them in a language laboratory, form a close system. The editor of the teaching material decides, once and for all, which patterns and what words should be drilled. Our approach is somehow different. To begin with, we associate patterns with goals, hence, the user can choose a pattern either by starting from the goal (the most natural way in a normal communicative setting) or from the structure (this might make sense in the case of rehearsing, i.e. learning). Next, the user can instantiate at every session a template (chosen directly by the user or indirectly via a goal) with the words that he would like to learn. This is something that was impossible in the past due to hardware constraints (books, tapes are closed systems), but with the advent of computers this is not a problem anymore at all.

5. Building and using the generator

As mentioned before, we suggest an open system, that is, a system that can be extended. Yet, there is a price to be paid: the user has to build the base from which he'd like the system to generate exercises from. Still, given the benefits, the price to be paid is small. The exercises generated are now meaningful for the student, in the sense that the system presents only words and structures the student wants to learn (words and structures he hasn't memorized yet). In addition, sentence order can be altered (random order, fixed order), whereas in traditional pattern drills sentence order is fixed once and for all.

The base is built interactively. For example, one specifies a goal, which will yield one or more patterns, which are then to be instantiated with words. Once this is done, the system has all the data needed to generate a set of sentences/exercises.

Let's take an example. Suppose you wanted to learn in Japanese how to ask for the location of an object. This goal could be expressed either via an abstract meta-language ((goal : place (object)) or by choosing from a menu. The result (system's answer) would be a template⁴ that the user is supposed to instantiate with words.

Suppose the decision were to instantiate *x* with any of the following words (bank, hospital and telephone), which could be done in the mother tongue, since the dictionary knows all the corresponding words (ginkoo, byooin, denwa). The chosen words can now play the role of a stimulus, the expectation being that the user integrates them in the right place of the template (figure 5).

SYSTEM (stimulus)	STUDENT	SYSTEM (confirmation)
bank	student's answer	<u>Ginkoo</u> wa doko desu ka?
hospital	student's answer	<u>Byooin</u> wa doko desu ka?
telephone	student's answer	<u>Denwa</u> wa doko desu ka?

Figure 5 : Training session

Figure 6 resumes again this kind of dialogue between the user and the machine. The program presents the words the student has decided to use (stimulus), to insert them then into the pattern to be learnt. The fact that the system displays its response allows the student to compare it with his own result. Hence it allows the learning, correction or strengthening of the response.

User's input	
(goal)	PLACE <OBJECT> ? [where is <OBJECT>?]
<OBJECT>	taxi stand
System's output : Takushii-noriba wa doko desu ka?	
System's stimulus: bank	

Figure 6 : Goal triggered template instantiated with a word

6. Discussion

The method presented here allows not only to generate sentences, but also to generate exercises containing the

4 Here, « *x wa doko desu ka* », where *x* is the object to be specified, object whose location you'd like to find out about, *wa* is the topic marker, and the remainder (*doko desu ka*) is roughly the English equivalent to the question « where is ».

words and syntactic structures the user wants learn. What makes this approach interesting is the fact that our tool can accommodate to the users' needs which may vary indefinitely. In addition, unlike with traditional patterns, our patterns are tightly coupled to communicative goals. In this respect they correspond well to a normal communicative situation where the starting point is an intention rather than the direct choice of a linguistic resource (pattern, sentence structure).

It should be noted that the building of this kind of generator is fairly straightforward in as the major part of the information needed is in the dictionary. Basically we need a set of templates, their corresponding goals and a morphological component. In order to prevent ill-formed sentences one could use subcategorization features, information which is also in the dictionary.

Computers are a medium escaping many of the constraints tapes or books are prone to. They allow for variable order of presentation, dynamic updating of words and much more. By and large they accommodate admirably well to the needs of the user, provided that the designer allows them to.

7. Conclusion

We started from the premise that one can do many more things than expected by enhancing an existing electronic dictionary. Our emphasis has been on sentence generation. While the examples given were all specific to learning Japanese, the approach taken is generic. Hence the approach should also hold for other languages. While pattern drills are not a panacea per se, they are quite useful if used properly and at the right moment.

8. References

- [1] Kant E (1781) Critique of pure reason, translated by Max Müller, Anchor Books, Garden City, NY
- [2] Goffman E (1974) Frame analysis. An Essay on the Organization of Experience. Harper & Row, Cambridge, Mass.
- [3] Harris Z (1951) Methods in Structural Linguistics, Chicago, University of Chicago Press
- [4] Fries C (1952) The Structure of English. An Introduction to the Construction of English Sentences. New York : Harcourt , Brace
- [5] Roberts P (1962) English Sentences, Harcourt Brace & World, New York
- [6] van Dijk T (1977) Semantic Macro-Structures and Knowledge frames in Discourse Comprehension. In M.A. Just & P.A.Carpenter (Eds.), Cognitive Processes in Comprehension, Hillsdale, N.J., 3-31
- [7] Mc Keown K. (1984) Text Generation, Cambridge University Press, Cambridge, U.K.
- [8] Bartlett F. (1932) Remembering, Cambridge, Cambridge University Press
- [9] Koffka K (1947) The Principles of Gestalt Psychology. New York, Harcourt Brace & World

- [10] Piaget J (1970) Piaget's Theory. In P.Mussen (Ed.) Carmichael's Manual of Child Psychology, vol. 1, New York, Wiley, 318-323
- [11] Bruner J (1973) Beyond the Information Given: Studies in the Psychology of Knowing. New York, W. Norton
- [12] Rumelhart D (1975) Notes on a Schema for Stories. In D.G.Bobrow & A.M.Collins (Eds.), Representation and Understanding : Studies in Cognitive Science. New York, Academic Press, 211-236
- [13] Mandler J (1979) Categorical and schematic organization in memory. In Puff,C. (Ed.), Memory organization and structure, N.Y., Academic Press, 259-299
- [14] Ausubel D (1980) Schemata, Cognitive Structure and Advance Organizers: a Reply to Anderson, Spiro and Anderson. American Educational Research Journal, 17, 3
- [15] Minsky M (1975) A Framework for Representing Knowledge. In Winston, P. (Ed.) The Psychology of Computer Vision. New York, Mc Graw Hill, 211-277
- [16] Wilks Y (1975) A Preferential Pattern-Seeking Semantics for Natural Language Inference. Artificial Intelligence, 6:1, 53-74
- [17] Schank R and Abelson R (1977) Scripts, Plans, Goals and Understanding : an Inquiry into Human Knowledge Structures. Hillsdale, N.J.
- [18] Lado R (1964) Language Teaching: a Scientific Approach, Mc Graw Hill, New York
- [19] Rivers W (1972) Speaking in many tongues: essays in foreign language teaching, Newbury House, Rowley
- [20] Skinner B (1957) Verbal Behavior. New York : Appleton-Century -Crofts
- [21] Chomsky N (1959) Review of Verbal Behavior, by B.F. Skinners. Language 35, 26-58