System Demonstration

ITSVOX

Eric Wehrli
LATL - Dept. of Linguistics
University of Geneva

1 Introduction

The huge increase in computer-based information exchange as well as the convergence of telecommunication and information technology have lead to an important demand for tools capable of processing documents in multimedia/multilingual environments. Examples of such tools include (i) systems which can convert information from one medium to another (ie. text —> speech), (ii) systems which can translate information from one language to another (ie. German —> English), or (iii) system which can convert a text in one language into an acoustic representation in another language (ie. written English —> spoken German).

ITSVox is a prototype system under development at LATL which addresses capabilities (i)-(iii) above. Specifically, ITSVox is a rnultilingual/multimodal processing system which combines speech synthesis and interactive translation technologies to produce a multilingual text-to-speech system, that is a system able to "read" a document either in its original language or in another language. The translation engine is based on the ITS-2 interactive model *(cf.* Ramluckun & Wehrli, 1993, Wehrli, 1994). As for the speech output components, they are based on our GB-parsers *(cf.* Laenzlinger & Wehrli, 1991, Wehrli, 1992) for the linguistic aspects, along with a prosodic component developed by the LAIP-University of Lausanne. The actual signal generation is provided by the MBRPsola system for the French synthesizer and by the DecTalk system for English.

ITSVox runs under MSWindows and can handle MSWord documents. On-going developments include (i) a WWW version, which will be able to process HTML documents, (ii) the addition of German, first as an additional source language and later as a target language with speech output, and (iii) the addition of a speech input component (to be developed in collaboration with IDIAP) in order to achieve speech to speech translation.

The next two sections discuss some of the properties of ITSVox, namely its interactive translation engine, and its general architecture, which takes advantage of the larger-than-expected similarities in the linguistic treatment underlying translation and speech synthesis.

2 Interactive translation

Perhaps the main problem in MT is that "superficial" systems (ie. older systems, which only perform partial and shallow linguistic analyses) provide results which do not meet most users' expectations, while "deeper" systems (ie. systems which try to perform a complete syntactic analysis, along with elements of semantics), which in principle could yield much better and much

more reliable translations, lack robustness and tend to be overwhelmed by ambiguities to such an extent that they cannot be applied to "real" texts[1].

The interactive approach to translation offers an attractive solution to the conflict between robustness and quality. Underlying the interactive model is the basic idea that a translation system needs to have access to the kind of knowledge and reasoning capabilities which is so natural for human beings (but hopelessly hard for computers!) in order to solve many ambiguity problems. An interactive translation system may consult its user when it faces problems it cannot solve by itself[2]. Perhaps the most interesting development in connection with interactive translation has been the shift of focus, from translators to monolingual authors. Underlying this new trend is the idea that all the interaction could be made in the source language, in which case knowledge of the target language would not necessary[3]. Such a system is primarily designed for authors (rather than translators) who want to produce in some other language(s). This line of research is currently pursued at the University of Grenoble *(cf.* Boitet, 1990; Blanchon, 1994), as well as at the LATL.

2.1 Interaction in source language

ITSVox is interactive in the sense that it can request on-line information from the user. Typically, interaction takes the form of clarification dialogues. Furthermore, all interactions are conducted in source language only, which means that target knowledge is not a prerequisite for users of ITSVox. User consultation can occur at several levels of the translation process. First, at the lexicographic level, if an input sentence contains unknown words. In such cases, the system opens an editing window with the input sentence and asks the user to correct or modify the sentence.

At the syntactic level, interaction occurs when the parser faces difficult ambiguities, for instance when the resolution of an ambiguity depends on contextual or extra-linguistic knowledge, as in the case of some prepositional phrase attachments or coordination structures. By far, the most frequent cases of interaction occur during transfer, to a large extent due to the fact that lexical correspondences are all too often of the many-to-many variety, even at the abstract level of lexemes. It is also at this level that our decision to restrict dialogues to the source language is the most challenging. While some cases of polysemy can be disambiguated relatively easily for instance on the basis of a gender distinction in the source sentence, as in (1), other cases such as the (much simplified) ones in (2)-(3) are obviously much harder to handle, unless additional information is included in the bilingual dictionary.

(l)a. Jean regarde les voiles.
'Jean is looking at the sails/veils'

b. voiles:

[1] Sophisticated approaches have been investigated by several research groups over the last 20 years, but none has lead to large-scale prototypes, let alone commercial systems.

[2] The idea of combining man and machine to achieve the task of translation is by no means new. It was first applied by Kaplan and Kay in the *Mind* system *(cf.* Kay, 1973), and a little later by A. Melby *(cf.* Melby 1987). These earlier systems were abandoned for various reasons, including the fact that they requested too much assistance *(cf.* Blanchon, 1994 for a detailed review of these systems). A new generation of interactive systems, based on much more powerful linguistic engines (and of course much more suitable hardware) has been developed since the late 1980s, for instance the N-Tran system *(cf.* Jones and Tsuji, 1990).

[3] As usual in the translation field, revision of the translation is made by a target language specialist, who will make the necessary stylistic modifications.

masculin (le voile)
féminin (la voile)

(2)a. Jean n'aime pas les avocats.
'Jean doesn't like lawyers/advocadoes'

b. avocats:
homme de loi *(lawyer)*
fruit *(fruit)*

Another common case of interaction that occurs during transfer concerns the interpretation of pronouns, or rather the determination of their antecedent. In an sentence such as (3), the possessive son could refer either to *Jean,* to *Marie* or (less likely) to some other person, depending on contexts.

(3) Jean dit a Marie que son livre se vend bien.
'Jean told Marie that his/her book is selling well'

In such a case, a dialogue box specifying all possible (SL) antecedents is presented to the user, who can select the most appropriate one(s).

3 Translation with speech output

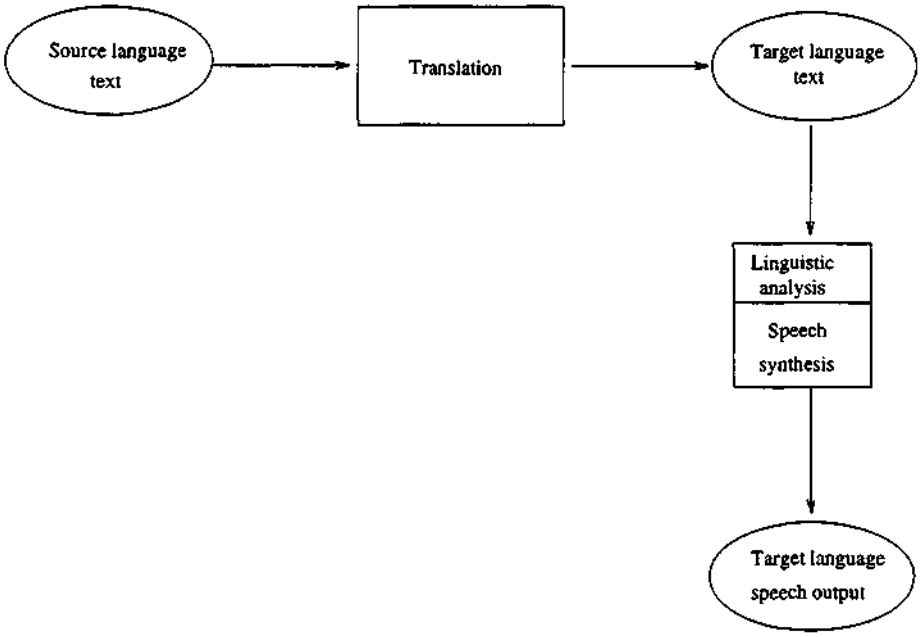
In the fast growing field of translation with speech output, there is usually no direct relation between the translation module and the speech synthesis module. Both components are considered as black boxes, the first one producing a target language text, while the second one is a standard target language text-to-speech system. This architecture is illustrated in (4) :

Such an organization lead to a significant amount of reduplication, in particular with respect to some of the most difficult aspects of natural language processing such as the disambiguation of lexical items, or the determination of phrases and their attachment types. Reduplication comes from the fact that all the abstract linguistic knowledge computed during the transfer and generation phases of the translation are confined to the translation black box. As a consequence, they are not available for further processing, which means that the speech system must recompute the linguistic knowledge necessary for speech synthesis (a task achieved by the linguistic analyzer in the diagram (4)).

Good quality speech synthesis systems must carry out a significant amount of linguistic analysis in order (i) to disambiguate homographs which are not homophones (words with the same spelling but different pronunciations such as *to lead/the lead, to wind/the wind, he read/to read, he records/the records, etc.,* (ii) to derive the syntactic structure which is used to segment sentences into phrases, to set accent levels, *etc.,* and finally to determine an appropriate prosodic pattern. In a language like French, the type of attachment is crucial to determine whether a liaison between a word ending with a (latent) consonant and a word starting with a vowel is obligatory/possible/impossible[4].

[4] For instance, liaison is obligatory between a prenominal adjective and a noun *(e.g. petit animal),* or between a determiner and a noun (e.g. *les amis),* or between a pronominal subject and a verb *(e.g. ils arrivent).* Liaison is optional between an auxiliary verb and a main verb *(e.g. il est arrivé)* and impossible between a non-pronominal subject and a verb *(e.g. les animaux ont soif).*

(4) Structure of the black box model

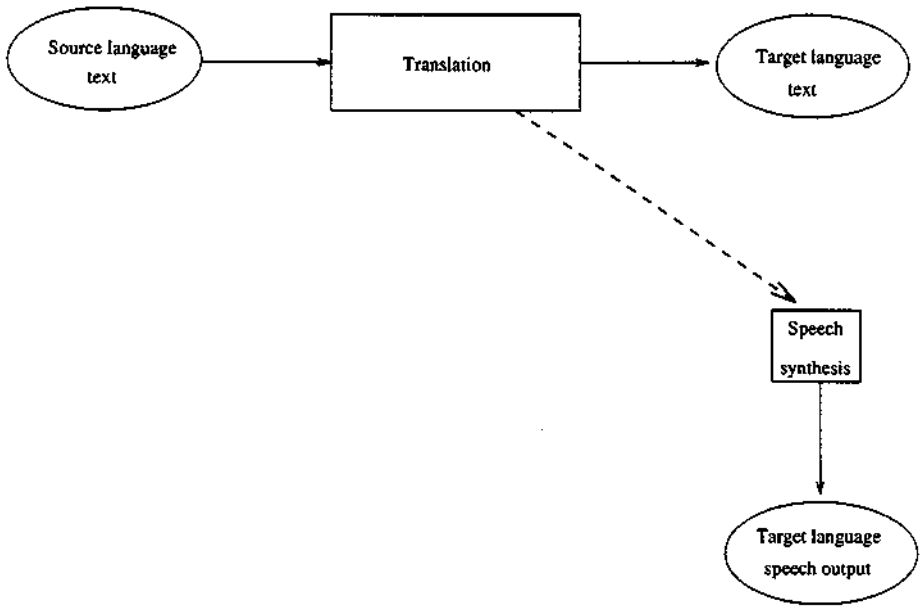


Such information is available during the translation process. High-level translation transfer modules map abstract representations of source language sentences into equivalent abstract representations for the target language, which undergo a series of (mostly) syntactic transformations to derive the so-called surface structure representations. Finally the surface structures undergo a morphological process, to determine the graphemic form of each word. A target sentence as outputed by the translation system is simply the concatenation of all the lexical leaves of the surface structure representation of that sentence *(cf.* Wehrli, 1994 for details about the translation module, and more generally about the ITS-2 interactive translation model).

It turns out that in a linguistically-sound machine translation system, the surface structure representations specify all the lexical, morphological and syntactic information that a speech synthesis system needs. Therefore, we can consider a more direct mapping between the translation module and the speech module. Specifically, in the ITSVox system the translation module can output abstract linguistic representations (enriched phrase-structure trees). Such structures can be passed directly to the speech component. This new architecture, dubbed "integrated model", is depicted in (5) :

Compared with the black box organization given in figure 1, the integrated model avoids redundancy by establishing a direct link between an internal state of the translation module and the speech synthesis module. Specifically, the linguistic structures computed by the transfer and generation module of the translation component are no longer lost, as they are in the black box model. Rather, they can be exported and passed directly to the speech synthesis system, which no longer needs to perform a linguistic analysis. As expected, the benefits of this model are (i) a much better quality of the output (fewer pronunciation errors and more appropriate prosody), and (ii) a significant increased in efficiency (no reduplication).

(5) Structure of the integrated model



References

Blanchon, H. 1994. *Lidia-1 : Une première maquette vers la TA interactive "pour tous"*, doctoral thesis, Université Joseph Fourier - Grenoble 1, GETA.

Boitet, C. 1990. "Towards Personal MT: on some aspects of the LIDIA project", Proceeding of *COLING90,* Helsinki, 30-35.

Jones, D. and J. Tsujii, 1990. "Interactive High-quality Machine Translation for Monolinguals", *Proceedings of the 3rd Conference on Methodological Issues in Machine Translation of Natural Language,* University of Texas, Austin.

Kay, M. 1973. "The MIND System", in Courant Computer Science Symposium 8, *Natural Language Processing,* New York, Algorithmics Press, 155-188.

Laenzlinger, C. and E. Wehrli, 1991. "FIPS : Un analyseur interactif pour le français", *TA Informations*, 32:2, 35-49.

Melby, A. 1987. "On Human-Machine Interaction in Translation" in S. Nirenburg (ed.) *Machine Translation: Theoretical and Methodological Issues*, Cambridge, Cambridge University Press.

Ramluckun, M. et E. Wehrli (1993). "ITS-2 : an interactive personal translation system" *Actes du collogue de l'EACL*, 476-477.

Wehrli, E. 1992. "The IPS system", in C. Boitet (ed.) *COLING-92,* 870-874.

Wehrli, E. 1994. "Traduction interactive : problemes et solutions (?)", in A. Clas et P. Bouillon (ed.), TA-TAO : *Recherches de pointe et applications immédiates*, Montréal, Aupelf-Uref, 333-342.