COUPLING AN AUTOMATIC DICTATION SYSTEM WITH A GRAMMAR CHECKER

Jean-Pierre CHANOD, Marc EL-BEZE, Sylvie GUILLEMIN-LANNE
IBM France, Paris Scientific Center

Automatic dictation systems (ADS) are nowadays powerful and reliable. However, some inadequacies of the underlying models still cause errors. In this paper, we are essentially interested in the language model implemented in the linguistic component, and we leave aside the acoustic module. More precisely, we aim at improving this linguistic model by coupling the ADS with a syntactic parser, able to diagnose and correct grammatical errors. We describe the characteristics of such a coupling, and show how the performance of the ADS improves with the actual coupling realized for French between the Tangora ADS and the grammar checker developed at the IBM France Scientific Center.

Description of the Tangora system

The Tangora system is implemented on a personal computer IBM PS/2 or IBM RS/6000. A vocal I/O card is added, as well as a specialized card equipped with two micro-processors, which provide the needed power for the decoding algorithms. The programs are written in assembly or C.

The multi-lingual aspect of the Tangora system (DeGennaro 91) constitutes a major asset. Indeed, it was initially conceived for English (Averbuch, 87) by the F. Jelinek team (IBM T. J. Watson Research Center), but it was adapted since to process Italian, German and French inputs. As a whole, the average error rate is close to 5%. But problems specific to each language require adapted solutions.

The user is required to train the system by uttering 100 sentences during an enrollment phase, and to manage slight pauses between two words. For the French system, liaisons at this time are prohibited.

Architecture of the system

The voice signal is submitted to a chain of signal processing, in order to extract acoustic parameters from the sound wave. Thus, the data flow is reduced from 30,000 to 100 bytes per second. Two passes of acoustic evaluation are performed: a relatively gross pass (so-called Fast Match) selects a first list of candidate words (around 500 words); this list is further reduced thanks to the language model (see below), so that only a small number of remaining candidates are submitted to a second, more precise, acoustic pass (so-called Detailed Match). Storage constraints as well as the methods used to provide the language model explain that the size of the dictionary is limited to about 20,000 entries.

The decoding algorithm

This algorithm determines the more likely uttered sequence of words. It works from left to right by combining the various scores estimated by the acoustic and linguistic models, according to a so-called stack decoding strategy. At this stage, the elementary operation consists in expanding the best existing hypothesis which is not yet expanded, i. e. it consists in keeping the sentence segment, which, followed by the contemplated current word, is rated with the highest likelihood.

Methods

If one formulates the problem of speech recognition according to an information theory approach, one naturally chooses probabilistic models among all available language models (Jelinek, 76). The trigram (Cerf, 90), triPOS[1] (Derouault, 84), or trilemma (Derouault, 90) models offer ways of estimating the probability of any sequence of words. For instance, formula of the trigram model:

$$P(W_1^n) = P(w_1) \times P(w_2/w_1) \times \prod_{i=3}^{n} P(w_i/w_{i-2}, w_{i-1})$$

The analysis of decoding errors show that half of them are due to the acoustic model, the other half being associated with the

[1] Model based on triplets of parts of speech (POS).

language model. Actually, the number of homophones being quite high (2.6) In an inflected language such as French, it is clear that no acoustic model, as perfect as it may be, can produce a satisfactory decoding without the support of a language model.

Power and limitations of probabilistic language models

Probabilistic language models are powerful enough to considerably reduce ambiguities that the acoustic model alone cannot solve. However, they suffer from punctual imperfections that are bound to their formulation. This is clearly shown by testing a probabilistic model on the lattice formed by the set of the homophones of the words of every sentence. The decoding obtained by searching for the maximum likelihood path (Cerf, 91) gives an error rate close to 3%, thus showing some of the inadequacies of the probabilistic language models.

Besides, and again for reliability reasons, statistics need to be gathered from large learning corpora (tens or even hundreds of millions words). In spite of all the preliminary cleaning that may be done (automatic correction of typos, tripled consonants for instance), such a huge corpus contains a certain number of grammatical errors, that introduce noise in the model.

Probabilistic estimations are produced by counting triplets of words or grammatical classes. In any of the trigram, triPOS or trilemma models, a word is generally predicted according to the two preceding words, classes or lemmas only. However, grammatical rules may apply to larger frames. Not only the rules often apply to words located out of the window used by the probabilistic model, but also grammatically significant words are to be found either in previous or in posterior position. Let us mention, as illustrations, some phenomena for which the probabilistic model does not fit:

- Appositions and interpolated clauses increase the distance between elements which must agree:
  Plusieurs PARTIS d'opposition de gauche, notamment le parti communiste, PARTAGENT ce point de vue.

- Predicting a word thanks to the preceding words does not allow the system to appropriately control person agreement when the subject follows the verb. Example:
  Que sont DEVENUS les principaux PROTAGONISTES de la victoire du onze novembre?

- Moreover, some confusions due to homophony induce changes of grammatical category, that require a complete interpretation of the sentence to be properly diagnosed, as in "et"/"est" (conjunction/verb) or "à"/"a" (preposition/verb).

Coupling the ADS with the grammar checker

To bring a solution to the problems described above, we propose to perform a grammatical analysis after the decoding operation. The grammatical analysis applies to the best of the hypotheses selected by the ADS. It serves as a basis to diagnose grammatical errors and to suggest corrections[2] .

The syntactic parser must prove powerful and reliable enough to effectively improve the performance of the ADS. It must provide a broad coverage, in order to cope with a large variety of texts, the source and the domain of which are not known in advance. It must also compute a global analysis of the sentence in order to fill the deficiencies of the probabilistic model.

Description of the syntactic parser

The syntactic parser we use meets the requirements described above (Chanod 91). It is actually conceived to provide the global syntactic analysis of extremely diversified texts.

It is based on an original linguistic strategy developed by Karen Jensen for US English (Heidorn 82, Jensen, 86). The parser initially

[2] A similar approach was tested in English, but only to detect grammatically incorrect sentences (Bellegarda 92)

computes a syntactic sketch, which represents the likeliest syntactic surface structure of the sentence; at this stage, such phenomena as coordinations, ellipses, interpolated clauses, if not totally resolved, do not block the parsing. The analysis is based on the so-called *relaxed approach*, which consists in rejecting linguistic constraints which, as pertinent as they may be in descriptive linguistics, are rarely satisfied *stricto sensu* in the surface structures of free texts. This strategy proves to broaden the coverage of the grammar as well as it allows the parser to deal with erroneous texts.

Architecture of the parser:

The system is written in PLNLP (Programming Language for Natural Language Processing, G. Heidorn, 72). It includes:

- A morphologic dictionary (50,000 lemmas plus their inflection tables),[1]
- A morpho-syntactic dictionary, which describes the sub-categorizations attached to each lemma,
- A set of more than 300 PLNLP production rules, which produce the syntactic sketches,
- A set of procedures built to re-interpret the syntactic sketches and to diagnose errors,
- A form generator, which provides corrected forms.

Indeed, some other techniques are also used. Strong syntactic constraints are relaxed during a second pass; it allows the system to detect errors which induce major syntactic changes (for instance confusion "et/est"), while forbidding undesired or too numerous parses. Fitted parses are computed in case the global analysis fails (Jensen, 83) and multiple parses are ranked thanks to specific procedures (Heidorn, 76). This last point allows the system to automatically select the strongest hypothesis, according to the linguistic features (including the grammar errors) of the syntactic trees.

Adaptation of the parser to the ADS

As mentioned above, many grammatical errors in written French are actually caused by homophones (gender, number agreement, confusion between infinitive and past participle, "chantez/chanter", "et/est", etc.). The parser, initially built for written French, is thus well prepared to detect errors produced by an ADS.

It can however be adapted to the specific needs of the ADS, by adding specific procedures (detection of ill-recognized frozen phrases, etc.), and by filtering out non-homophonic corrections, or corrections which do not belong to the list of candidates initially proposed by the ADS.

Indeed, post-processing procedures are largely used to diagnose errors after the syntactic tree has been computed. This offers the immense advantage of making the system evolutionary: it can be easily modified, in order to improve the scope of the detections. This made the adaptation of the grammar checker to the ADS quite straightforward.

Description of the processing chain

In case of the ADS, the coupling is done by a simple call to the parser for each sentence. In case of the homophone scheme, the diagram of the processing chain is shown in the following figure:

[1] These 50,000 lemmas produce about 350,000 inflected forms, which largely exceeds the 20,000 forms used by the Tangora system.

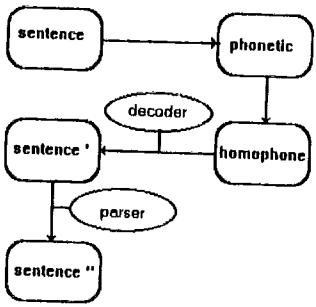

Figure 1. Coupling Diagram

Experiences

Our tests were carried on the following texts:

corp1	AFP dispatches	(1000 words)
corp2	AFP dispatches	(3221 words)
corp3	e-mail notes	(1909 words)
corp4	grammar books	(1337 words)

Only the CORP1 file was obtained through a real decoding; the other corpora were processed by automatically generating their homophones.

Results

The experiments were made at an early stage of the coupling. They could certainly be improved with more extensive tests, as the adaptation of the grammar checker to the ADS would gain in accuracy.

Percentage of erroneous words left uncorrected

	LM without parser	with parser
corp1	4.5%	3.6%
corp2	4.6%	3.6%
corp3	6.3%	6.1%[4]
corp4	7%	5.8%

Given the high performance of the ADS and the difficulty to improve it in the frame of the probabilistic model, the improvement of around 1% observed on three of the test corpora is very promising.

Samples of corrected sentences:

Example 1: Subject-predicate, attributive adjective-noun, subject-verb agreement

Les conditions sont très durs mais le pays, devenue indéfendable, les acceptent.

After parsing, the suggested correction is:

Les conditions sont très DURES mais le pays, DEVENU indéfendable, les ACCEPTE.

Example 2: subject-verb agreement; confusion between the conjunction "et" and the verbal form "est" :

Le fait que le héros de chacun des trois romans soient différents et révélateurs.

After parsing, the suggested correction is:

Le fait que le héros de chacun des trois romans SOIT DIFFéRENT EST RéVéLATEUR.

Example 3: Confusion between the verbal form "a" and the preposition "à"; Confusion between the past participle and the infinitive form of the corresponding verb.

Ce document est a faire signé recto et verso par le propriétaire et par le gestionnaire.

After parsing, the suggested correction is:

Ce document est à faire SIGNER recto et verso par le propriétaire et par le gestionnaire.

Conclusion

Coupling the ADS and the syntactic parser meets the initially assigned objectives quite satisfactorily: broad coverage of the texts parsed by the grammar, meaningful percentage of justified corrections, adequacy of the syntactic parser to the types of errors specifically generated by the decoder.

The tests that we performed on various corpora are all the more encouraging, since a great deal of the remaining errors result from semantic ambiguities that no grammar checker based upon a syntactic analysis of the sentence can detect.

[4] **The bad results of the CORP3 file are due in great part to the difficulties of e-mail, that make parsing less accurate.**

L'âge de la MER le plus fréquent à l'accouchement est de vingt-six ans .

A subsidiary advantage of the coupling would be to detect errors that would not be produced by the ADS but by the speaker him/herself (punctuation, stylistic infelicities, mood of subordinate clauses, etc.). Not only we may contemplate transcribing as accurately as possible the words of a speaker, but also offering him/her a stylistic aid.

References

Averbuch A. et al., 1987: Experiments with the TANGORA 20,000 word Speech Recognizer, *Proceedings of ICASSP*, Dallas, pp. 701-704.

Bellegarda J., Braden-Harder L., Jensen K., Kanevsky D., Zadrozny W., 1992: "Post-recognizer language processing: applications to speech, handwriting", submitted to EUSIPCO'92.

Cerf-Danon H., de La Noue P., Diringer L., El-Bèze M., Marcadet J.C., 1990: "A 20,000 words, automatic speech recognizer. Adaptation to French of the US TANGORA system", Nato 1990.

Cerf-Danon H., El-Bèze M., 1991: "Three different Probabilistic Language Models: Comparison and Combination", ICASSP 1991.

Chanod J-P., 1991: Analyse automatique d'erreurs: stratégie linguistique et computationnelle, Colloque informatique et Langue naturelle, 23-24 janvier 91, Llana Univ. de Nantes.

DeGennaro S., Cerf-Danon H., Ferretti M., Gonzales J., Keppel E., 1991: "Tangora - a large vocabulary speech recognition system for five languages", EuroSpeech 1991, Genoa.

Derouault A-M., Mérialdo B., 1984: "Language modeling at the syntactic level" 7th International Conference on Pattern Recognition, August 1984, Montreal.

Derouault A-M., El-Bèze M., 1990: "A Morphological Model for Large Vocabulary Speech Recognition", ICASSP 1990.

Heidorn, G.E., 1972: *Natural Language Inputs to a Simulation Programming System*, Ph.D. dissertation, Yale University.

Heidorn G.E., Jensen K., Miller L.A., Byrd R.J., Chodorow M.S., 1982: "The EPISTLE Text-Critiquing System", *IBM system journal, vol.21, n°3.*

Heidorn, G.E., 1976: "An Easily Computed Metric for Ranking Alternative Parses", Presented at the Fourteenth Annual Meeting of the ACL, San Francisco, October 1976.

Jelinek F., 1976: "Continuous Speech Recognition by Statistical Methods", *Proceedings of the IEEE, Vol 64*, April 1976.

Jensen, K., Heidorn, G.E., 1983: "The Fitted Parse: 100% Parsing Capability in a Syntactic Grammar of English", *Proc. Conf. on Applied Natural Language Processing*, Santa Monica, California, pp.93-98.

Jensen, K. 1986: "A Broad-Coverage Computational Syntax of English", *Unpublished documents*, IBM T.J. Watson Research Center, Yorktown Heights, N.Y.