

Stentor, a new Computer-Aided Transcription software for French language

Thierry Spriet

SténoMédia,
20 Bd Bastille, 75012 Paris, France
<http://www.stenomedia.com>

thierry.spriet@stenomedia.com

Abstract

We are presenting in this paper the technology used in STENTOR, the new software for Computer-Aided Transcription (CAT) in French.

The stenotypy domain is between speech recognition and text analysis.

In the most used stenotypy method in France, words are described by their sounds using a syllabic approach.

This implies some difficulties because of the high number of homophones in French: this problem is similar with a classical speech recognition problem. Homophonic rate in the French language is about 1.8 and in the most widely used method of stenotypy in France is not adapted to reduce it.

On the other hand, most of time, we have at our disposal punctuation information and the use a comprehensive dictionary, which is more like texts analysis.

In STENTOR, in order to avoid homophone ambiguity, we are proposing a classical treatment based on the linear interpolation of a 3-class and a 3-gram statistical language models. Some adjustments were proposed, as a word-class factorization to reduce the linguistic model size.

The speeches which have to be processed are various and often highly specialized. Even with the use of a comprehensive dictionary, very often new words have to be introduced during the transcription process.

We use a specific training corpus about 4.5 million words issued from more than several hundreds hours of transcripts.

The rate of error of STENTOR was tested on a corpus of only 5 000 words, but the comparison test have shown that we are very competitive indeed.

Index Terms: stenotypy transcription, french homophony

1. Introduction

Stenotypy is a very good lab to apply and to develop researches in linguistic engineering. Between speech recognition and texts analysis, stenotypy transcription has to deal with words described by their pronunciation, but without acoustic treatment problems.

Applied to the French language, we have the same homophony problematic than in speech recognition. On the other hand, as in a text analysis approach, we have to work with very large vocabulary, with the

possibility of managing different dictionaries and adding news words during the transcription process.

In this paper, we first briefly compare stenotypy and speech recognition system. Then we explain some specific problem encountered in French language and their effect in CAT system. We present the technology used in Stentor, a new CAT system developed in order to process French language.

At last, we present some results from experiment made in order to evaluate the performance of this new system.

2. Stenotypy vs speech recognition

Computer-Aided Transcription (CAT) and automatic speech recognition (ASR) can be compared in two ways:

- the first one is about their use: why to use CAT systems instead of ASR system?
- the second one is about the similarities between the linguistic technologies used in both systems.

2.1. Using stenotypy

Why to continue to use stenotypy while speech recognition has improved a lot and has now a very small error rate? In fact, stenotypy is used in situations where some constraints can be resolved by speech recognition.

A great difference in stenotypy is the human interpretation made by the verbatim reporter. Even if this extra intervention has to be minimized, it allows to delete stammering and hesitation.

The stenotypist can also make a very powerful speaker identification even if two speakers speak together.

The stenotypist can also add some extra speech events like “*Mr. Smith leaves the room*” or “*Mr. Brown approves*”.

If we want to know what happens exactly in a meeting, during the examination of a witness or in a court room, we need this human intervention.

At last, ASR technology is not mature enough to efficiently process speech in a noisy environment, or overlapped speech: stenotypist's brain is still the best cognitive system to process such phenomena.

2.2. Stenotypy and speech recognition similarities

As in speech recognition, words in the most widely used French stenotypy method are described by their pronunciation.

The problematic is quite similar to speech recognition in French language, such as:

- acoustic variations due to typing errors,
- acoustic similarities due to ambiguities of the French stenotypy method used in France,
- high rate of homophones, plus ambiguities provided by the French stenotypy method,
- high rate of homographs, which cannot be efficiently reduced by a n-gram model;
- long span syntactic constraints, which are not well modeled by n-class model but need special models like in [1]

3. Text analysis context

Stenotypy is used in a lot of specific areas such as court reporting, technical meetings, boards of directors, conventions, arbitrations, conciliation boards and so on. Each time we need a specific vocabulary depending on the firm or on the agenda of the meeting. Even when using a very large vocabulary, we have to manage with new words. To do that, we take in account the *unknown word* in the linguistic model and offer the possibility to the stenotypist to add words while working realtime.

Something also very interesting in computer-aided transcription is the knowledge of the sentence boundaries. A linguistic model including this information is much more efficient.

4. Professional and historic context

For several years now, French stenotypists use the same method in France. Step by step, some of them adapt this method to avoid ambiguities which generate errors when using a CAT system. To develop the Stentor software, we had to take this into account.

The problem is that these adaptations are personal and lead to specific dictionaries for each user.

Each change needs a more or less long period of adaptation for the stenotypist. It is impossible to ignore these modifications, and we have to deal with that.

When the modification is limited to the orthography a word to drop an ambiguity between this word and another one, it is easy to include this to a user dictionary.

But when the adaptation is about the grammatical word description, this can generate some problems with our linguistic models.

We have decided to be independent of the stenotypy method and to accept all user variations. But we recommend to use the real syntactic class of the words, in order to be coherent with the linguistic models used in the Stentor CAT system.

5. linguistic models

The figure 1 gives a good example of what it just be presented above. In this example, we have only 4 keystrokes:

```
POUL  
L  
F*E  
ST*OUD
```

As we can observe, we have between 4 and 10 candidates for each reference word. In this case, the solution is very easy for a human analysis, but the computer can find at least 3 good paths in this graph.

Except if we are in a meeting about avian influenza or something about chickens, the right path is "*pour lever ce doute*".

The problem can be more complex if in this meeting we have someone named *Mister Pourlever*. In this case, we need the context of the sentence to decide.

Without the use of language model, any path in this graph could be proposed by a CAT system: this shows the need, in French language, of the use of language models.

The software STENTOR actually use a mixed approach, using statistics and knowledge rules as presented in [2]. We use a linear combination of a 3-gram and a 3-class language models. This technology is classical for automatic speech recognition.

In order to reduce the size of the model, we apply some factorizations on the 3-class model.

5.1.3-gram model

The 3-gram model is in fact a combination of 3-gram, 2-gram en 1-gram models. We use a specific training

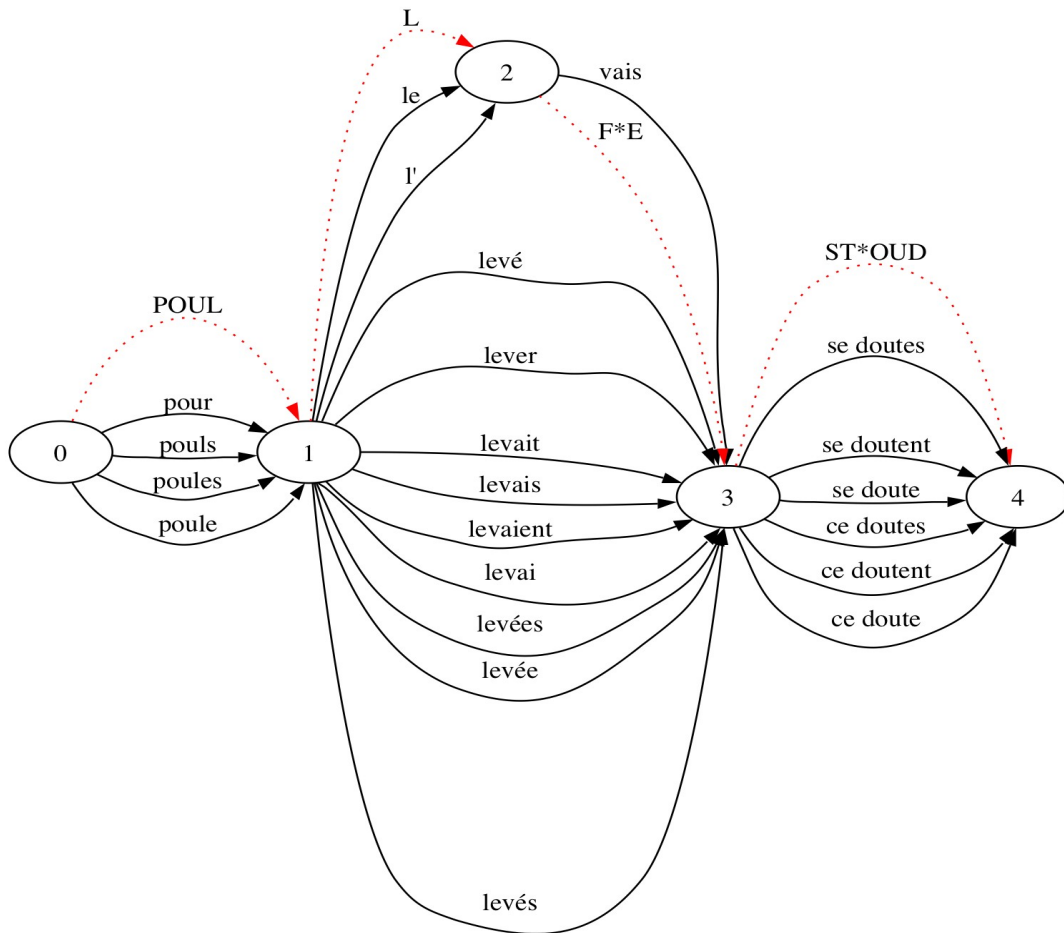


Figure 1: example of ambiguity graph in french stenotypy

corpus of 4.5 millions of words and a lexicon of about 150K most used words.

The training corpus is a corrected transcription of some 200 hours of meetings, boards of directors or town councils.

We have extracted the 150K most used words of this corpus to build the lexicon.

We have a special token for unknown words and the n-gram models are training with this token. So even if these models cannot guess an unknown word, they don't eliminate this possibility.

5.2. 3-class model

This model is based on statistics association of Part Of Speech (POS).

The training corpus was annotated with POS. This tagging was performed with a stochastic part of speech tagger [3] using a modified tag set of 105 POS. We add 2 special POS, one for foreign words and one for acronyms.

To reduce the size of the model, we have merged some classes which had the same behavior.

6. Experiment

To estimate the quality of the transcription we decided to use the word error rate as used to evaluate speech recognition systems. To this end we use the NIST Scoring Toolkit (SCTK) as used in NIST international evaluation campaigns [4]. The word error rate takes into account word substitutions, word deletions and word insertions.

Stentor is based on statistical language modeling. So, it is necessary to estimate the language models on a training corpus. As seen above, the training corpus contains 4.5 million words and is composed by a collection of real data provided by stenotypists.

The test corpus is distinct from the training corpus and was manually corrected. There is no longer typing

errors: they were manually corrected too in order to evaluate only the CAT system, not the stenotypist.

We had not time to set up a large test corpus, so we have only a 5K words corpus. This test corpus is extracted from real data built by a stenotypist.

In this corpus we have a word error rate of 6%.

We also made a comparative test with the French computed-aided transcription TASF+ [5]. On the same test corpus, our software had 10% less word error than the TASF+ software using the same user dictionaries.

7. Conclusions

The first version of STENTOR is now out and can be used for computer-aided transcription in realtime reporting or post treatments.

The word error rate is already competitive and we are planning in reducing it further. For this, we are going to use stochastic Finite State Automata in order to take in account long span dependencies as in [6]. Of course, we first have to integrate a larger corpus in the training step. We hope that the French stenotypy community is ready to help us in this task, giving us a lot of public and corrected transcriptions.

Even if STENTOR is a good lab to apply our researches, it is also a professional software which offers a lot of functionalities for a more efficient work :

- audio-sync, for post correction, when the user selects a word or a place in the stenotypy, he can hear directly this passage;
- dictionaries builder, a set of tools which make easy dictionary management like insertion of new words, importation of a former dictionary, merging of dictionaries and so on;
- realtime word insertion, during a realtime reporting, the user can add very easily a new word (a proper noun for example). This new word will be taken in account immediately by the system.
- computer assisted correction: for each word of the transcription, the software proposes all the alternatives ordered by decreasing pertinence.
- short cuts, the most frequent short cuts used in word processing softwares are implemented in STENTOR.

8. References

- [1] Frédéric Béchet and Alexis Nasr and Thierry Spriet and Renato de Mori, Large Span Statistical Language Models: Application to Homophone Disambiguation in Large Vocabulary Speech Recognition in French, Eurospeech 99, p 1763-1766, Budapest, Hongrie, 1999
- [2] Thierry Spriet and Marc El-Bèze, Introduction of Rules into a Stochastic Approach for Language Modelling, Computational Models of Speech Pattern Processing, NATO ASI Series F, vol. 169, ed. Keith Ponting, pp. 350-355, 1998
- [3] Thierry Spriet, Marc El-bèze *Etiquetage probabiliste et contraintes syntaxiques. TALN 95*, 1995
- [4] Jonathan G. Fiscus Nicolas Radde John S. Garofolo Audrey Le Jerome Ajot Christophe Laprun, *The Rich Transcription 2005*, Spring Meeting Recognition Evaluation, 2005
- [5] <http://www.stenotype-grandjean.com/>
- [6] Alexis Nasr and Yannick Estève and Frédéric Béchet and Thierry Spriet and Renato de Mori., A Language Model Combining N-grams and Stochastic Finite State Automata, Eurospeech99, Vol 5 p 2175-2178, Budapest, Hungary, 1999