TEXT ANALYSIS FOR THE BELL LABS FRENCH TEXT-TO-SPEECH SYSTEM

Evelyne Tzoukermann

Bell Labs – Lucent Technologies

ABSTRACT

The Bell Labs text-to-speech synthesis system for French is part of a multilingual effort for text-to-speech generation which covers eight languages, including English, Italian, Russian, German, Chinese, Spanish, Japanese, etc. The text analysis modules are composed of three main layers of analysis, built in a unified framework of finite-state transducers (Mohri & Sproat, 1996; Sproat, 1996). This paper focuses on text analysis components of the text-to-speech system for French (Tzoukermann, 1994).

1. Introduction and underlying approach

The text analysis component consists of four main parts: the morphological analysis module, the language models, the grapheme-to-phoneme conversion rules, and the prosodic module. The system is built in a pipeline architecture, the output of which feeds the subsequent synthesis modules. In this paper, we will be describing the first three modules.

The work described here follows the lines of generative phonology (Schane, 1968). Our system consists or an architecture which reflects the bidirectionality of the generative process. That is, whereas generative phonology takes underlying (phonemic) representations as the basis for rule applications, our system takes surface forms, performs morphological analysis, and through a set of ordered rules, derives the word pronunciation.

The originality of this work lies in the fact that we use weighted finite-state transducer technology to perform complete analysis in the French system. Moreover, the implementation not only accounts for most orthographic representations, such as numerals, abbreviations, dates, currencies, etc, but we also solve the hard questions of French liaison, "mute e", and "aspirated h" using refined intermediate representations either in the form of traces or in the form of archigraphemes.

2. Morphological analysis of surface forms

The morphological component contains transducers for both inflected and non-inflected words. Non-inflected words include about 2500 adverbs, 235 conjunctions, 90 prepositions, 100 pronouns, 13 articles, and 290 interjections.

2.1. Inflected words

The system is based on a French lexicon constructed from different sources (Boyer, 1993), as well as a dictionary of acronyms available on the Internet. Additional lemmas have been acquired from text corpora, primarily from newspapers articles, such as Le Monde (ECI, 1989 and 1990). The total number of lexical entries is close to 90,000.

Verbs, nouns, and adjectives undergo complex morphological changes in French and thus require a high level of morphological analysis (Tzoukermann & Jacquemin, 1997). The morphological system is stem-based and the general framework follows the one instantiated in (Tzoukermann & Liberman, 1990), where stems are pre-compiled before analysis. French verbs can have from one to six stems for a verbal inflection. For instance, the stems for the verb "vouloir" are:

/{reg_v74-1} : (veu{verb})/
/{reg_v74-2} : (voul{verb})/
/{reg_v74-3} : (voud{verb})/
/{reg_v74-4} : (veul{verb})/
/{reg_v74-5} : (veuill{verb})/

and some of the inflectional paradigms for this verb appear as follows:

Paradigm	{reg_v74-1}	
Suffix	{++}x	{1st}{sing}{pres}{ind}
Suffix	{++}t	{3rd}{sing}{pres}{ind}
Paradigm	{reg_v74-2}	
Suffix	{++}oir	{infinitive}
Suffix	{++}ons	{1st}{plur}{pres}{ind}
Suffix	{++}ez	{2nd}{plur}{pres}{ind}
Suffix	{++}ais	{1st}{sing}{imper}{ind}
Suffix	{++}ussiez	{2nd}{plur}{imper}{sub}
Suffix	{++}ues	{past}{part}{fem}{plur}
Paradigm	{reg_v74-3}	
Suffix	{++}ras	{2nd}{sing}{fut}{ind}
Suffix	{++}raient	{3rd}{plur}{pres}{cond}
Paradigm	{reg_v74-4}	
Suffix	{++}ent	{3rd}{plur}{pres}{ind}
Paradigm	{reg_v74-5}	
Suffix	{++}es	{2nd}{sing}{pres}{sub}

For instance, label {reg_v74-1} corresponds to the first subparadigm attached to the verb inflection, and the compiler expands each entry in the word list to its labelled paradigm, as shown in Figure 1.

In some cases, lemmas have zero suffixation, and the paradigm is marked by the empty string ε or {Eps} in the system. For instance, adjectives ending in "-ais" such as "versaillais" require the following inflections. Notice that the first line covers both singular and plural for the masculine form:

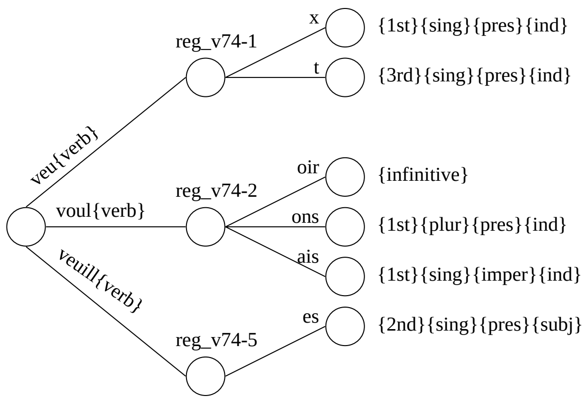

Figure 1: Transducer for some inflections for the verb *vouloir*.

Paradigm	{reg_adj2}	
Suffix	{++}{Eps}	{masc}{sing}{plur}
Suffix	{++}e	{fem}{sing}
Suffix	{++}es	{fem}{plur}

For the 11,400 verbs, 75 paradigms are necessary for verbal inflection, excluding irregular forms, such as "être" (to be), "avoir" (to have), and "aller" (to go). These 75 paradigms are composed of sub-paradigms, totalling 205 patterns. Interestingly, in comparison, only 18 paradigms are used for 59,000 nouns, and 30 paradigms for 22,200 adjectives. These numbers show clearly that the morphological complexity of French resides largely within the verbal system.

3. Handling other textual issues

A large part of the textual analysis is devoted to specific issues, such as acronyms and unknown words, numeral expansion, time and date, temperatures, currencies, etc. The length of this paper does not permit a full description of each of these questions, but all these specificities are implemented and correctly handled by the system.

4. Language models

This section deals with some of the core issues of French phonetic conversion. We show examples of how liaison and "mute e" are processed through traces preserving underlying representations. Language models encompass ambiguous pronunciations due to the position of the word in its given context. Syntactic knowledge is required and is implemented in the forms of local grammars. Once morphological analysis is achieved, we perform disambiguation on first homograph, and then agreement. The second part of the language models deals with all questions concerning liaison.

4.1. Homograph disambiguation

Homographs which are also homophones are disambiguated at this stage. For example, words such as "présidents" (presidents or [they] preside) /presidɑ̃/ (noun) or /presid/ (verb) can be disambiguated. This could easily be handled by a part-of-speech tagger (Tzoukermann & Radev, 1998) but as the number of heterophone homographs is small in French, the tagger was not used at this level. There are over ten rules to disambiguate all instances of nouns and verbs ending in "-ent". The following rule adds the feature {wrong} to the transition "les" (3rd-masc-fem-plural-clitic-pronoun) followed by a word boundary {##}, followed by any character ({Sigma} & !{##}) followed by a noun ending in "-ent":

{Eps} → {wrong} / __les{pron}{3rd}{masc}{fem}{plu}{clit}
{##}({Sigma} & !{##})*ent{noun};

Similar problem was found for "est" /ɛst/ (east) or /e/ (is). For disambiguation, we collected from corpora all the collocations in which the word "ouest" (west) occured and modeled "est" /ɛst/ in that fashion. In using this heuristic, we found that "est" /ɛst/ often followed prepositions, such as: "entre l'est" (between east), "vers l'est" (towards east), "dans l'est"(in the east), "de l'est" (from the east), " l'est" (at the east), etc, as well as expressions, such as "vent(s) d'est" (east wind(s)).

4.2. Disambiguation through agreement

Even though agreement has generally no direct impact on pronunciation, it is useful for removing ambiguities. For example, the noun phrase "son talent" (his/her talent) exhibits a number of ambiguities since "son" can also be a noun (bran) and "talent" can also be a verb (destroy) [1]. The following shows some possible analyses of "son talent" before disambiguation:

son{noun}{masc}{++}{sing}{##}
    tal{verb}{++}ent{3rd}{plur}{pres}{ind}{##}<1000.00>
son{noun}{masc}{++}{sing}{##}
    talent{noun}{masc}{++}{sing}{##}<1000.00>
son{adj}{poss}{masc}{sing}{clit}{##}
    tal{verb}{++}ent{3rd}{plur}{pres}{ind}{##}<1000.00>
son{adj}{poss}{masc}{sing}{clit}{##}
    talent{noun}{masc}{++}{sing}{##}<1000.00>

Transitions have weights associated with them and the lowest weight transition is the one selected. In the example, both [adj-noun] and [noun-verb] have similar weights. However, we need to rule out the [noun-verb] transition, and use the following agreement rule:

{sing} → {sing}<100.0> /
    __{clit}?{##} ({Sigma} & !({##}))*{plur};

The rule adds additional weight (<100.0>) to the {sing} feature when it is a clitic preceding any other word in the {plur} form. Thus, it will add weight to the [noun-verb] transition (<1100.00>) consequently favoring the [adj-noun] reading.

4.3. Liaison

Liaison is a complex phonological process (Schane, 1968; Tranel, 1990) resulting in surface phonetic alterations at

[1] Note that the verb "taler" has a rather obsolete usage. Nevertheless, since it exhibits a general problem (i.e. ruling out a [noun-verb] transition), it is treated as a general case.

the juncture of two words. It can occur inside phrase boundaries as well as across major phrase boundaries. In addition to the complexity of strictly linguistic processes, other factors influence liaison, including socio-cultural background and dialect preferences.

Liaison can be:

1. **Obligatory**. For example the words *un* (a) and *ami* (friend) when pronounced in isolation, are respectively [œ̃] and [ami]. Notice that the orthographic /n/ is not realized explicitly at the surface, but the vowel [œ̃] is nasalized. When the word *un* is an {article}{masc}-{sing} (and not a {pronoun}{masc}{sing}), it will undergo liaison with the following word, yielding the pronounciation [œ̃ **n**ami] with the liaison phoneme /n/ in bold.
2. **Optional**. The words *chez* (at) and *Annie* (Annie) in isolation are pronounced [ʃe] and [ani]. In succession, they can be pronounced either [ʃe ani] or [ʃe **z**ani] where the word-final orthographic *z* is fully realized by the insertion of the phoneme /z/ in bold.
3. **Forbidden**. In the phrase *j'en ai un aussi* (I have one also) pronounced [zɑ̃ nɛ œ̃ osi], liaison is blocked between *un* which is the {pronoun}{masc}{sing} and the adverb *aussi*. It would be considered an error to pronounced these two words [œ̃ nosi], although this type of error is typical of foreign speaker hypercorrection.

For our implementation, a set of intermediate symbols is used to represent liaison (Sproat *et al.*, 1997). For example, in the phrase "un premier enfant", the pronunciation of the words, if separated, would be [œ̃] [prəmje] [ɑ̃fɑ̃]. Within text analysis processing, the treatment will be as follows:

1. The output of morphological analysis is:
   un {##}un{art}{masc}{sing}{clit}{##}
   premier {##}pr{=e}mi{adj}{++}er{masc}{sing}{##}
   enfant {##}enfant{noun}{masc}{++}{sing}{##}
2. While the word ending is being processed, a trace $[L_{(V)C}]$ is appended at the end of each word: [œ̃$[L_n]$ prəmje$[L_r]$ ɑ̃fɑ̃$[L_t]$];
3. Liaison rules apply and in this context, state that:
   - liaison occurs between adjectives followed by nouns,
   - half closed unrouded front vowel [e] should substitute to its counterpart half open [ɛ],
   - orthographic consonant "r" should appear in the pronunciation.

   yielding the following realization: [œ̃ prəmjɛ r ɑ̃fɑ̃].

5. Grapheme-to-phoneme conversion

Grapheme-to-phoneme rules or letter-to-sound rules are applied to character strings with or without morphological analysis. As in phonology, we use typical left-to-right rules with context specification. The order of rule application matters; thus, suffixes and word endings take priority over the rest of the word. Verbal affixes and regular affixes are first applied; then, consonant rules followed by vowels rules.

5.1. Suffix rules

Verbal suffixes are handled separately due to their specific pronunciation. For instance, the form "attraperions" undergoes the following morphological analysis:

attrap{verb}{++}erions{1st}{plur}{pres}{cond}{##}
<500.00>

The typical rewriting rule for this verb will be:

erions → ərjɔ̃/ {verb}{++}
_({Person}{Grammatical}*({##}|{Boundary})) ;

where the suffix "-erions" is rewritten [ər j ɔ̃] when preceded by a {verb} feature and a morpheme boundary {++} and followed by a {Person} feature and any number of {Grammatical} features, such as {plur}{pres}{cond} in this case.

5.2. Consonant and vowel rules

The other rules, for consonant and vowels are then applied. For example, in order to disambiguate the graphemes "ien" and its pronunciations [jɛ], [jɑ̃], [jɛ̃], or [je], the following rules were used:

ien → jɛ/ _(n{LCons}?{Grammatical}?{fem}); /* parisienne/s */
ien → je / _(n) ; /* viennois */
ien → jɑ̃/ s _s|({Sigma} & !{Pos})* ; /* consciencieux */
ien → jɛ̃/ _({Cons}|{Epss}) ; /* czarien */

5.3. Other rules

Once all graphemes are rewritten into phonemic strings, accent applies. Even though French words have lexical stress, the stress needs to be assigned. For doing so, phonemic strings need to be properly syllabified so that accent apply on the last syllable. Additionally, functions words are not assigned accent. This data goes to an intonation module which interprets this information.

5.4. Archigraphemes and lexicalization

When grapheme-to-phoneme conversion cannot be handled by general-purpose pronunciation rules, lexicalization is used. In this context, this means that an archigrapheme replaces the character at the level of the lexical entry. This is how the French schwa or "mute e" and "aspirated h" are handled [2]. For the schwa, which is the only vowel in French which can be either fully realized, reduced, or completely removed, several rules were applied to disambiguate the grapheme "e" into /ɛ/, /e/, or /ə/. For example, the grapheme "-ess-" can be realized [əs], [ɛs], or

[2] Note that from a phonological point of view, it is possible to rule the French e-mute in observing the type of syllable (open or closed). Since syllables are not annotated as open or closed in the system and since the grapheme to phoneme rules apply only at a surface level, the schwa is mostly handled by lexicalization.

[es], in "dessous" (under), "accessibilité" (accessibility), and "abbesse" (abbesse) respectively, and some of the following rules were applied:

e → e / (pr)|(gr) _(ss) ; /* supresser */
e → ɛ/ (dr) _(ss) ; /* adressage */
e → ə/ r _|(ss) ; /* ressembler */

For the rest of the cases, we lexicalize the entries in the dictionary. The following examples show words extracted from the lexicon where the archigrapheme {=e} will be directly converted into the "mute e":

/{reg_adj} : (cliqu{=e}tant{adj}<500>)/
/{reg_adj} : (craqu{=e}tant{adj}<500>)/
/{reg_adj} : (ch{=e}v{=e}lu{adj}<500>)/
/{reg_adj8} : (s{=e}mestriel{adj}<500>)/
/{reg_nf} : (avant-premier{noun}{masc}<500>)/

In this way, the word "chevelu" (long-haired) will be correctly converted into /ʃəvəly/.

6. Related work

Extensive work has been done in the area of automatic grapheme-to-phoneme conversion for French. In the work of ((Divay & Guyomard, 1977)), a language named TOP was established for describing context-sensitive rules for phonetic transcription. Later (Divay & Vitale, 1997) presented a system of letter-to-sound for English and French. As in the Bell Labs system, the French system uses left to right rewriting rules with context specifications; it uses about 600 rules and 100 classes.

In (Yvon, 1996), self-learning techniques are used to convert French orthography. The system is based on the idea of pronuncing words by analogy to known lexical items. It offers interesting performance, but requires further additions so that the results of the chunck-based model apply better to a word-based model. The work of Boula de Mareuil (1997) goes in the direction of our approach in that the pronunciation is also built on a set of rules. The work handles general and particular problems of French conversion rules, including liaison, "mute e", disjunctive h, numeral expansion, etc. The system has been tested into the LIMSI synthetizer. Work at ICP (Aubergé, 1991; Belrhali, 1995) are also broad projects for working systems based on 60,000 dictionary entries and 2100 rules.

The CNET system with the work of (Larreur *et al.*, 1989) presents rewriting rules for French along with linguistic analysis to feed the intonation module. Linguistic analysis is tightly combined with a rather small dictionary (about 13,000 words) and the programs interpret the linguistic knowledge encoded in a set of tables.

Like most of the cited work, our system is a rule-based system covering proper names, numeral expansion, acronyms, date, time, temperature, and so on. The system consists of about 600 rules. The novel contribution lies in the fact that rules are compiled into weighted finite-state transducers, allowing some transitions to be selected over others. Additionally, we use intermediate representations to keep track of underlying phenomena such as liaison.

7. Results and conclusion

The system is fully implemented as part of the Bell Labs text-to-speech system for French. Two kinds of evaluation have been pursued, including perceptual test by listening to the most 100 frequent sentences extracted from large corpora through a greedy algorithm, and listening to random French text selected on the World Wide Web. The overall performance is near 99% accuracy. We are in the process of setting up rigourous tests over different kinds of corpora in order to measure performance on a range of parameters.

The system is publicly available and can be heard and tested on: http://www.bell-labs.com/project/tts/french.html

References

Aubergé, Véronique. 1991. *La synthèse de la parole: 'des règles au lexique'.* Ph.D. thesis, Université de Grenoble, Grenoble, France.

Belrhali, R. 1995. *Phonétisation automatique d'un lexique général du français; systémique et émergence linguistique.* Ph.D. thesis, Université de Grenoble, Grenoble, France.

Boyer, Martin. 1993. *Dictionnaire du français.* Québec, Canada: Hydro-Québec, GNU General Public License.

Divay, Michel, & Guyomard, Marc. 1977. *Conception et réalisation sur ordinateur d'un programme de transcritpion graphémo-phonétique du français.* Ph.D. thesis, Université de Rennes, Rennes, France.

Divay, Michel, & Vitale, Anthony J. 1997. Algorithms for Grapheme-to-Phoneme Translation for English and French: Applications. *Computational Linguistics*, **23**(4), 495–523.

ECI. 1989 and 1990. European Corpus Initiative, Le Monde Newspaper.

Larreur, D., Emerard, F., & Marty, F. 1989. Linguistic and Prosodic Processing for a Text-to-Speech Synthesis System. *In: Eurospeech.* European Conference on Speech Communication and Technology, Paris, France.

Mohri, Mehryar, & Sproat, Richard. 1996. An Efficient Compiler for Weighted Rewrite Rules. *Pages 231–238 of: 34th Annual Meeting of the Association for Computational Linguistics.*

Schane, Sanford A. 1968. *French phonology and morphology.* Cambridge, USA: MIT Press.

Sproat, Richard. 1996. Multilingual Text Analysis for Text-to-Speech Synthesis. *In: Proceedings of the ECAI-96 Workshop on Extended Finite State Models of Language.* European Conference on Artificial Intelligence, Budapest, Hungary.

Sproat, Richard, Möbius, Bernd, Maeda, Kazuaki, & Tzoukermann, Evelyne. 1997. Multilingual Text Analysis. *Chap. 3, pages 31–87 of:* Sproat, Richard (ed), *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach.* Boston, MA: Kluwer.

Tranel, Bernard. 1990. On suppletion and French liaison. *Probus*, **2**, 169–208.

Tzoukermann, Evelyne. 1994. Text-to-Speech for French. Proceedings of the ESCA Workshop on Speech Synthesis, Mohonk, USA.

Tzoukermann, Evelyne, & Jacquemin, Christian. 1997. Analyse Automatique de la Morphologie Dérivationnelle et filtrage de mots possibles. *Pages 251–260 of: Workshop on Mots possibles et mots existants (Possible Words and Occurring Words), Silexicales*, vol. 1. Lille, France: SILEX, University of Lille III.

Tzoukermann, Evelyne, & Liberman, Mark Y. 1990. A Finite-State Morphological Processor for Spanish. *In: Proceedings of the Thirteenth International Conference on Computational Linguistics.* International Conference on Computational Linguistics, Helsinki, Finland.

Tzoukermann, Evelyne, & Radev, Dragomir R. 1998. *Use of Weighted Finite State Transducers in Part of Speech Tagging.* Cambridge University Press.

Yvon, François. 1996 (September). Grapheme-to-Phoneme Conversion using Multiple Unbounded Overlapping Chunks. Proceedings of NeMLaP-II, Bilkent University, Ankara, Turkey.