LINGUISTIC CONTRIBUTIONS TO TEXT-TO-SPEECH COMPUTER PROGRAMS FOR FRENCH

Pierre TRESCASES
University of Western Ontario, Canada

Matthew CROCKER
University of British Columbia, Canada

Abstract

This paper presents the development of computer programs used for transcribing French text into phonetic speech. Based on an earlier program (Maggs & Trescases, 1980) made of a set of some 900 text-to-phonetics rules and of a compact set of prosodic rules for synthetic speech, the present research was primarily aiming at developing the best possible algorithm to account for practically any word in the French general lexicon, as opposed to most frequently used words only. In order to considerably enhance the original rules, these were systematically tested against a 50,000 word French pronunciation dictionary and an equally important corpus of texts which were all entered in an IBM PC-XT. At the same time, a set of syntactic rules was developed for most liaisons, mandatory and forbidden, and homographs to be found in French speech. The result is a set of some 4,000 conversion rules. Tested against the 50,000 words of the pronunciation dictionary, they yield a low percentage of error. Errors in text are similarly minimal, and due mainly to foreign (English) words. To allow faster programing, while accounting for most commonly used words in French, a compact set of 2,000 rules has been developed. It is essentially based on a statistical analysis yielding most frequently used general rules. The aim of both algorithms has been to make possible better text-to-speech software for French.

1. Introduction

The transcription programs operate by pattern matching the source text with the set of text-to-speech rules.

The system is actually a three pass one. The first, the primary objective of the research,converts any text using the traditional spelling system into phonetic symbols using the International Phonetic Alphabet (IPA). A rule has thus the general form:

context$_L$ [text] context$_R$ = phonetics;

This reads: in a given context, if any, indicated at left and/or right of brackets, the written form involved in the rule yields the phonetic representation appearing in the second part of the equation. For instance,

th[ai+]=aajj; (cf. Appendix B)

means that the written form aï preceded by th (itself preceded by a space indicating that it must be the start of a lexeme), and no matter what follows (since there is no space before the equal sign, making the rule applicable not only to thaï but to Taïwan, taïwanais, etc.) should be pronounced /aj/ (which, in our ad-hoc phonetic system, see Appendix A, i represented by ai+).

The second system converts the phonetic representation into a more specific one, which determines the appropriate duration and inflection of each phoneme. The third converts the more specific representation into actual codes or mnemonics which are required by a specific speech synthesizer. Each phase uses precisely the same transcription mechanisms. That is, the source text is transcribed, then the basic phonetics are transcribed, then the specific phonetics are transcribed - all using the same procedure. The only change is the set of rules being used for each phrase.

Our initial basic program written in 1979-80 with Peter B. Maggs at the University of Illinois at Urbana, and acquired in 1983 by the Catholic School Board of Montreal (CÉCM) for use in a secondary school for the handicapped, consisted of less than a thousand text-to-phonetics and speech synthesis rules - that is less than the thoroughly researched "maximum" program of 1,158 rules of Nina Catach (1984). Ours had been tested against the 5,000 word Juilland Frequency Dictionary of French Words (1970), and an equally small corpus of texts (cf. Maggs & Trescases, 1980). It was therefore felt that the conversion rules, and, above all, the treatment of liaison and homographs could greatly be improved upon, in order to yield greater accuracy in the phonetic transcription of non technical texts. This was made possible in the Fall of 1984 after receiving two major grands from the Social Sciences and Humanities Research Council (SSHRC) of Canada, and further assistance from the University of New Brunswick.

2. Main objective of the research

The research started in the Fall of 1984 had a a triple objective.

First, it sought to develop a more complex set of phonetization rules for common, standard French, and not for a specific variety of French, such as the one from France of Quebec. The rules are therefore based on the phonological system of "standard" French, including the nasal vowel /œ̃/ and the borrowed English phoneme /ŋ/ but using only front /a/ to transcribe the two a's, and /nj/ to transcribe the dorso-palatal nasal /ɲ/. Geminates within words were generally not accounted for.

The main objective was to develop the best possible set of rules for transcription of any word in texts based on the general lexicon of French. To this effect rules should yield a correct transcription for most frequent cases of homographs, and most commonly used "exceptions" and rare words.

The algorithm thus developed was intended for efficient use in micro-computers. In an effort to achieve accurate transcription, one

had to be careful not to overextend the program. A balance had continually to be struck between writing rules that cannot possibly be finite, if aimed at exhausting all the possibilities of the lexicon and of speech, and program manageability and efficiency. The task of attempting to take into account new exceptions to the rules, and all possible contexts for homographs and liaisons, without mentioning English words and abbreviations, is virtually never-ending.

3. Approach

Various constraints influenced the approach taken.

3.1 Constraints

The approach followed was contingent on the various constraints under which research had to be conducted (over a three-year period by a computer programmer and the main investigator, with a full teaching load). Research had therefore to be based on the original program developed four years earlier. An overhaul of the set of rules taking into account syllabification and the graphemic system of French as set forth by N. Catach (1980) was not, for all practical purposes, possible.

Consequently, if the above-described phonological system chosen for representing common French pronunciation little differs from the one used by N. Catach (1984), our graphic system is not always based on graphemes, now considered as the units in this system. As a result, for instance, we have rules for the graphic forms aill and ail instead of rules for the graphemes ill and il preceded by a. This, and the lack of syllabification, is no doubt a cause of some lesser efficiency in the text to phonetics correspondences. Whereas N. Catach's program consists of 240 basic grapheme-phoneme rules, ours is made up of some 325 such rules, directly evolved from the original basic program.

However, as said earlier, our primary objective in the present research was to correctly transcribe the maximum number of words belonging to the general lexicon while dealing with homographs and liaison in discourse in a more sophisticated way.

3.2 Tools

In order to greatly improve upon the original program, rules had to be tested against large corpuses.

The first was made up of the French Pronunciation Dictionary by A. Martinet and H. Walter (1973) based on actual usage, and not on a prescriptive norm. The dictionary is made up of some 50,000 words, out of which 10,000, chosen for the investigation, register variations in pronunciation. A computer program was written for checking the pronunciation of each word, as given by the program rules, against the first - or only - and any other pronunciation registered in the dictionary (cf. Appendix C). Statistics of matches against first, second... tenth pronunciation were simultaneously computed, as shown below for all words starting with letter a:

```
----------*> File Statistics
Total words in file     :4322
Total words checked     :4322
Total words listed      :525
Total number of matches:4022 %=93.06
==========> Breakdown of matches :
Matches in pron # 1 are:3630 %=83.99
Matches in pron # 2 are: 372 %= 8.61
Matches in pron # 3 are:  14 %= 0.32
Matches in pron # 4 are:   2 %= 0.05
Matches in pron # 5 are:   2 %= 0.05
Matches in pron # 6 are:   0 %= 0.00
Matches in pron # 7 are:   2 %= 0.05
Matches in pron # 8 are:   0 %= 0.00
Matches in pron # 9 are:   0 %= 0.00
Matches in pron #10 are:   0 %= 0.00

----------*> Neutral File Statistics
Total words in file     :4322
Total words checked     :4322
Total words listed      :525
Total number of matches:4274 %=98.89
==========> Breakdown of matches :
Matches in pron # 1 are:4039 %=93.45
Matches in pron # 2 are: 225 %= 5.21
Matches in pron # 3 are:   6 %= 0.14
Matches in pron # 4 are:   4 %= 0.09
Matches in pron # 5 are:   0 %= 0.00
Matches in pron # 6 are:   0 %= 0.00
Matches in pron # 7 are:   0 %= 0.00
Matches in pron # 8 are:   0 %= 0.00
Matches in pron # 9 are:   0 %= 0.00
Matches in pron #10 are:   0 %= 0.00
```

As a result, non matches as recorded on print-outs allowed for systematic correction of faulty rules, and, consequently, continuous improvements to the algorithm. It was moreover possible, when a word was mispronounced in the dictionary or in a text, to detect which rule, or ordering, was involved. The following example shows how all the rules involved in the pronunciation of a word could be listed:

```
[ ]=
[ ]=
[p]=pp
[e]CC=eh
[ch]=sh
[b]=bb
[l]=ll
 pechbl[en]de =ehnn
[d]=dd
[e]-=
[ ]=
```

French: pechblende IPA: ppehshbbllehnndd

As numerous non matches - not considered as genuine pronunciation errors but rather as possible variations - resulted from transcriptions with open instead of closed unstressed, vowels, or vice-versa, in the case of E, O and EU, the neutralization applying to A (selection of one phoneme, the front one) was extended for correction's sake. Real pronunciation errors were then made to stand-out in the print-outs, and facilitated systematic correction of errors in the rules. As far as the graphic-to-phonetics rules themselves are concerned, to the exclusion of other types of rules (for liaisons, homographs, etc...), correction stopped when it was felt that the rules produced an optimal phonetic transcription (cf. Appendix D) considering the system used. Given the fact that there are many permitted

variations even within standard pronunciation, in particular in unstressed syllables, compromises had to be often reached in the development of the rules.

It should not, however, be construed from the use of Martinet & Walter's dictionary that it was used as a model of "good" French pronunciation to be reproduced. It was basically used, as mentioned earlier above, as a tool for systematic corrections of the rules, as well as a reference dictionary registering variations in pronunciation. In the development of the conversion rules between graphic forms and their phonetic representation, constant use was made of other reference works, mainly of Le Petit Robert (PR) and the Dictionary of Pronunciation of A. Lerond (Larousse, 1980). In fact no effort at all was made to reproduce some features to be found in the general pronunciation revealed by the data such as the pronunciation of geminates and the use of /õe/vs/$\tilde{\varepsilon}$/.

The second corpus, made up of texts, as opposed to lists of single words, also consisted of some 50,000 words. These were mainly articles found in French and Quebec general information magazines. They were not only intended as a further way of checking on graphic-phonetic rules but, above all, as a means of enhancing rules for liaisons and homographs.

We will now briefly discuss two major areas on which the research focused.

4. Treatment of liaisons

Improvements to the earlier basic program proceeded along two lines: a) increasing the number of words causing obligatory, and to a lesser extent semi-obligatory, liaisons; b) refining on rules taking more liaison contexts into account without creating forbidden liaisons at the same time.

4.1 Selection of words making up liaison rules

Here again, a balance had to be struck between the will to take as many liaison, contexts into account, and the over extension of the rules, not to mention that absolutely all liaison possibilities cannot be allowed for selection or words causing liaison - to be added to lists in original program based on Juilland frequency list - was further made a) according to frequencies given in the Basic Orthographic Lists or LOB (Catach, 1984); b) linguistic awareness, since it was no longer a matter of focusing on a 5,000 but on a 50,000 word lexicon. Memory limitations on the program contributed moreover all along to reduce the list of liaison rules as much as possible. To this effect, semi-obligatory liaisons were generally only entered for verbs, and the decision was finally taken because it was felt they made comprehension of synthetic speech easier.

The program includes close to 300 - mostly obligatory - liaison rules, out of which some 180 for adjectives alone. Here is an example of such a rule (cf. Appendix B)

```
vil[ain ]L=ehnn;
```

It reads in a liaison context L the masculine adjective ending is pronounced like the feminine.

In the earlier programs, all liaison rules looked like the above. Liaison was always applied, except before a consonant or so called aspirated h words. These have to be listed, and liaison is prevented before them through the use of a macro symbol referring to their list. However, applying liaison to all other vowel initial words is bound to cause unwanted liaisons such as in:

Il est vilain et réussit

4.2 Use of macros

It was consequently necessary to introduce constraints on the context of application of liaison rules, and this was accomplished through the use of two different macro symbols. The first, E, prevents application of the liaison rule only before a consonant initial or an aspirated h word. The second, L, restrict liaison before entire parts of speech. In this case before prepositions (P), relative and interrogative pronouns (R), conjunctions (J) and verbs (B). These complex macros - as opposed to simple ones such as V for vowel and C for consonant - therefore require lists of words to be available to the system. To increase efficiency, such lists had to be drawn and ordered in terms of the frequency of the words to be included.

By preventing liaison before vowel initial words belonging to specific parts of speech represented by other macro symbols, and included in lists available to the system, the use of the macro symbol L in liaison rules allows for a correct transcription, without liaison, of premier, usually a determiner, in such a context as follows:

Le premier/et mon second se ressemblent.

It must, however, be obvious that the treatment of ambiguities, in the case of liaison or homographs, through the immediate context of the word involved in the rule, has its limitations. Besides the fact that all parts of speech are not represented, macros and the lists they make available to the system are far from being exhaustive.

4.3 Use of hard vs. soft hyphen

The presence of an hyphen after, or before a word, or part of a word, to which a liaison rule applies, is bound to modify the nature of the liaison. For instance, if final s in états can be pronounced /z/ in des États unis par des liens économiques, liaison must be made in les États-unis. Similarly a third person singular verbal ending in d should be pronounced /t/ then followed by an hyphen and a personal pronoun (inversion of the subject) whereas liaison is optional or forbidden in other cases.

For such reasons, two different types of hyphen were introduced. The presence of a hard (H) in a rule will cause the rule to apply only if there is a hyphen in the text. The rule:

[ed]HV=eett;

will make the final ed be pronounced /et/ in a liaison context such as in sied-il or pied-à-terre, and only in such cases of mandatory hyphen.

Use of hard hyphen has thus several advantages as it restricts the application of a rule, in liaison as well as other cases, and therefore reduces the overall size of the program. It proved particularly useful in the case of compounds. For instance, it allows for generalization of liaison involving the final consonant of the first element of a compound, as in:

[t]H=tt; (pot-au feu; and also dit-il...)

Any such generalization will of course entail listing rules accounting for exceptions (chat-huant) and will cause some new compounds to be incorrectly pronounced (restaurant-artisan).

On the other hand, the use of a soft hyphen (-) in a rule will allow the rule to be applied whether the word is or not preceded and/or followed by an hyphen. The rule

-mar[s] =ss;

will cause pronunciation of the final s whether the word appears on its own, is found in a compound (champ-de-mars) or in a phrase such as: en février-mars. The use of soft hyphen helps solve the problem of the unsystematic use of hyphen in French compounds, thus reducing pronunciation errors in computer processed texts.

5. Treatment of homographs

As above mentioned, use of macro symbols, referring to special lists of words available to the system, makes possible generalization concerning the context surrounding graphic forms appearing in rules. They were particularly useful as new symbols were created to this effect, in the treatment of ambiguities, which constitute a major problem to be solved in developing text-to-phonetics rules for French. There are numerous cases of homographs in French involving various problems (le bus vs. tu bus; nous inventions des inventions, etc...). The most complex and frequent one no doubt involves the -ent final not pronounced in verbal endings but otherwise pronounced /a/. This case will be used to illustrate our general treatment of ambiguity in pronunciation, as summed up in the following chart:

TREATMENT OF -ENT

CATEGORIES

I Rules for adjectives causing liaison

ex.: appar(ent) L=/at/;

II Other exceptions IN -ENT

ex.: abriv(ent)=/a/;

III Rules of desambiguisation (in terms of graphic forms of endings)

ex.: em(ent)=/a/;

IV Syntaxic rules

ex.: I W(ent)=; (ils affluent)

V Homographs

ex.: S afflu(ent)=/a/;
(un affluent vs. ils affluent)

VI General rule: (ent)=;

In the order of application of rules, Part I is made up of most frequent -ent adjectives involved in liaisons. Part II included all frequent words whose ending is pronounced and cannot be derived from general rules (around 200). Part III is made up of general rules existing in the earlier program, and taking into account the immediate context before the -ent final. For example, all words ending in -ement have their final pronounced. Part VI constitutes the most general rule, whereby the final is not pronounced. Parts IV and V were additions to the program, and make abundant use of macro symbols in an effort to reduce ambiguity by defining the context beyond the immediate graphemes preceding the form on which the rule bears.

For instance, the rule

I W[ent]=; (ils affluent)

where I can be any personal pronoun and W is any lexeme, or part of it, preceded by a space, cause any final -ent, in a word immediately preceded by a personal pronoun, not to be pronounced.

Similarly, the rule

T A W Ws W[ent] =; (les très heureux parents rayonnent de joie)

makes use of T, for plural determiners, and A, which can be any adverb from a special list or nothing.

The program includes some 100 rules of the kind, which evolved from the analysis of the texts. They were tested against these and such nonsense sentences as this one, correctly transcribed by the program: Les parents du président président en occident le serpent et serpent le froment de l'opulent président. Naturally the same rules that produce a correct pronunciation in the above sentence, could produce errors in others. It is obvious again that this treatment of ambiguity is not error proof for several reasons: 1) more remote contexts then the ones that can be defined by

such rules are sometimes involved; 2) not all parts of speech, in particular adjectives, are represented by a macro symbol; 3) special lists made available to the system are not exhaustive; 4) in the present state of the program, some symbols could not indifferently be used right and left of the brackets in the rules, thus restricting generalization; 5) finally, it will take a sophisticated semantic and syntactic text analysis to interpret certain cases of ambiguity (ils ont tous leurs livres) or sociolinguistic variation (plus). Only the use of huge corpuses of texts will test the correctness of the rules developed for dealing with homographs in French. Time and memory limitations prevented any further development of the rules, which appear to be adequate for all practical purposes, and in the present state of the research.

6. Statistics and results

In the absence, and impossibility, of one norm of French pronunciation, there can be no absolutely rigorous method of evaluating the correctness of graphic to phonetics rules, especially in the discourse. The dictionary of pronunciation entered into the computer in order to test our rules has nevertheless allowed for an assessment of the rules developed as applied to a lexicon.

6.1 Assessment of rule correctness

The pronunciation of each headword by the rules could be matched against any pronunciation registered in the dictionary, and words not matching first pronunciation were listed, as shown in the treatment of the first page of the dictionary (cf. Appendix C). Matches, with or without neutralization of double timbre vowels, could then be computed as shown in the statistics for letter A (cf. Appendix D). Statistics for the entire dictionary indicate a 90.9% and 96.2% match average respectively with or without neutralization. Matches against the first pronunciation only average respectively 81.6% and 89.7%. It should be noted that a proportion of errors is caused by treatment of some graphic signs used in dictionary (hyphens and apostrophes found in compounds etc...) and that our use of /nj/for/ɲ/, /œ̃/ for some of the dictionary's /ɛ̃/ and the fact that the rules did not generally attempted to produce geminates likewise failed to produce matches. The sole purpose of the above-mentioned statistics is to help give a general idea of the performance of the rules as applied to a pronunciation dictionary.

6.2 Development of a compact set of rules

Statistics were computed on the application of all the rules to both corpuses of 50,000 words each - lexicon and texts. For instance, rules applied more than 10,000 times or so are the following:

32 248	a=a;	25 515	i=i;
31 990	t=t;	18 822	r=r;
26 229 (final)	e =;	17 325	d=d;
25 891	l=l;	17 068	p=p;
16 948	m=m;	13 915	n=n;
15 725	(o)C= ;	12 478	é=e;
14 382	s=s;	10 031	u=y;
14 121	c=k;	9 457	h=h;

These statistics could also be used to determine the frequency of each grapheme. They indicate, for instance, that the frequency for a, à and â has been respectively 97%, 2% and less than 1% compared to 92%, 7% and 1% as quoted by N. Catach (1980).

The ordering of rules was also confirmed by the statistics, as more general rules should be applied more often than rules listed before in the program.

Finally, during the somewhat limited time allowed for the research, statistics on frequency of application of the rules were mainly exploited to develop a more compact program for faster use with a speech synthesizer. Without going into particulars here, the least applied rules, that is under ten times or so, were eliminated, unless it was felt that the low frequency was mainly due to the unsufficient size of the corpus of text (made of some 50,000 words). We therefore turned to a shorter dictionary of some 16,000 words, Le Nouveau Larousse des débutants, Édition canadienne (1981) for help in selecting little used rules to be kept. The selection process resulted in a "compact" set of 2,000 rules, half the size of the "complete" program. Both sets include some 200 rules for English words. They give an acceptable pronunciation for roughly 2/3 of English words likely to be found in a French dictionary. It is estimated that around 350 rules are needed for the 500 non integrated English borrowings registered in the dictionary used. This represents a ratio of 2/3 of a rule for every English word taken into account! With regards to length of the rules and efficiency of the program, English words, as well as proper nouns and abbreviations (O.N.V.), constitute complex problems still to be solved.

Finally, statistics on the application of rules to corpuses could be directly used to completely reorder the rules according to frequency, and therefore greatly increase the efficiency of the program. This, and other matters touched upon, would be sufficient to prove that, although far less complex to develop than rules for speech synthesis itself, text-to-phonetics rules for French will remain a good field.

References

Catach, N., L'Orthographe française, Paris, Nathan, 1980

_____, Les listes orthographiques de base, Paris, Nathan, 1984.

_____, La phonétisation automatique du français, Paris, CNRS, 1984.

Lerond, A., Dictionnaire de la prononciation, Paris, Larousse, 1980.

Martinet, A. et Walter, H., Dictionnaire de la prononciation française dans son usage réel, Paris, France-Expansion, 1973.

Trescases, P. et Maggs, P.B., "De l'écrit à l'oral. Un programme sur ordinateur pour machine à parler à l'usage des aveugles francophones", Le français moderne, juillet 1980, n° 3, pp. 224-244.

Le Petit Robert, 1978.

Nouveau Larousse des débutants, édition canadienne, 1981.

APPENDIX A

TABLE OF SYMBOLS

Phonetic	Dictionary/Computer	Program Output
	@	aa
	@~	an
	<~	eh
	<~	en
	>	oh
	>"	oh
	&~	on
	Y~	oe
	a"	un
	a°	ui
	O-	ae
	%	th
	#	ew
	$	gn
	+	ng
	=	sh
	_	zh
	^	eu
a	a	aa
b	b	bb
.	.	.
.	.	.
z	z	zz

The following denote how the accents are transcribed when text is input:

Acute	<
Grave	>
Circumflex	*
Diaerisis	+
Cedilla	%

APPENDIX C

LIST ERRORS ON 1ST PRONUNCIATION

```
a    IPA:  aa
@ (abcjlmpvwx) - ah NEUT Match
a (dgknrty) - aa Mach Neut Match

an-  IPA:  an
an . aann

abaisser  IPA:  aabbeessee
ab<se (cdgjlpvw) - aabbehssee NEUT Match
abese (bkmnrtxy) - aabbeessee Match NEUT Match

abasourdissant  IPA:  aabbaazzuurrddiissan
abasurdis@~ (ackmnptvw) - aabbaassuurrddiissan
abazurdis@~(djlrxy) - aabbaazzuurrddiissan
Match NEUT Match
```

APPENDIX B

RULE LISTING

```
;
 [a] =aa;
[a>]=aa;
 ur[aeus] =eeyyss;
[ae<]=aaee;
[ae>]=aaeh;
 [ae]schne =eh;
 m[ae+]=aaeh;
 m[ae]=aaeh;
-p[ae]lla-=aaeh;
 t[ae]l =aaeh;
[ae]=ee;
[ai*]Cer =ee;
[ai*]Cez =ee;
[ai*]Ce<=ee;
[ai*]Ci=ee;
[ai*]=eh;
 dj[ai+n] =aaiinn;
 j[ai+n] =aaiinn;
[ai+]V=aajj;
 samour[ai+] =aajj;
-t[ai+]=aajj;
 th[ai+]=aajj;
[ai+]=aaii;
[aient]H=ehtt;
[aient ]E=ehtt;
[aient]=eh;
 s[aie]tt=ehjjeh;
 g[aie] =ee;
[aie]Ce<=ee;
[aie]Ci=ee;
 vr[aies ]L=eezz;
[aie]=eh;
[a*illi]V=aajj;
[aillie =aajjii;
 jo[ailli]e=aalljj;
[a*ill]V=aajj;
[aill]V=aajj;
[aill]V=ehll;
cockt[aill]=ehll;
[aill=aajj;
[ai]Cer =ee;
[ai]Cez =ee;
[ai]Ce<=ee;
[ai]Ci=ee;
[ai]Cu=ee;
[ai]gner =ee;
[ai]gnez =ee;
[ai]gne<=ee;
[ai]grier =ee;
[ai]guer =ee;
[ai]guez =ee;
[ai]gue<=ee;
[ai]gui=ee;
[ai]Ker =ee;
[ai]Kez =ee;
[ai]Ke<=ee;
[ai]Ki=ee;
[ai]Ku=ee;
[aim]V=ehmm;
[aim]=en;
[ain]V=ehnn;
 cert[ain ]L=ehnn;
 proch[ain ]L=ehnn;
 s[ain ]L=ehnn;
 soud[ain ]L=ehnn;
 v[ain ]L=ehnn;
 vil[ain ]L=ehnn;
[ain]=en;
[a*]=aa;
f[ai]sV=eu;
 ass[ai] =aajj;
 g[ai] =ee;
[ai]=eh;
[a]nV=aa;
[a]nn=aa;
-gr[ands-]L=anzz;
-[an]h=aann;
 cy[an]hydrique =aann;
 d[ans ]E=anzz;
 d[ans] =an;
 d[an] =aann;
  hetm[an] =aann;
  karm[an] =aann;
 ombudsm[an] =aann;
 p[an]helle<ni=aann;
 pir[an]ha =aann;
 s[an]he<drin=aann;
-[and]-=ehnndd;
-barm[an]-=aann;
-f[an]-=aann;
 gentlem[an] =aann;
-m[an]-=aann;
 recordm[an] =aann;
[an]=an;
[a]mV=aa;
[a]mm=aa;
 rh[amn]ace<e =aammnn;
[a]mn=aa;
[am]b=an;
```

APPENDIX D

```
l' europe un re*ve impossible ?
ll ewrrohpp un rrehvv enppohssibbll?.

non !
nnon..

l' europe existe , elle ne s' est j' amais aussi
bien porté<e !
ll ewrrohpp ehggzziisatt /. ehll nneu ss eh zh
aammehzzoossii bbjjen ppohrrttee..

des europe<ens convaincus de l' avenir de leur
continent ,
ddeezzewrrohppeeen kkonvvenkkyy ddeu ll
aavvnniirr ddeu lloerr kkonttiinnan /.

on en entend beaucoup .
/ohnnannnanttan bbookkuu..

mais parfois , le lieu d' ou>l' on parle importe
autant que ce qu' on a a> dire : bruno kreisky
/mmeh ppaarrffwwaa /. lleu lljjew dd uu ll oo
ppaarrll enppohrrtt /oottan kkeu sseu kk ohnnaa
/aa ddiirr /. bbrryynnoo kkrrehsskkii.

, l' ex-chancelier d' autriche il a pris sa
retraite l' an passé<,
/. ll ehkksshansseulljjee dd oottrriish /iill aa
pprrii ssaa rreuttrrehtt ll an ppaassee /.
```