

Étude de quelques problèmes de phonétisation dans un système de synthèse de la parole à partir de SMS

Rémi Bove¹

Équipe DELIC – Université de Provence
29, Av. Robert Schuman, 13621 Aix-en-Provence Cedex 1
remi.bove@voila.fr

Mots-clefs – Keywords

SMS, phonétisation, synthèse de la parole.

SMS, phonetisation, speech synthesis.

Résumé – Abstract

Cet article présente une étude dont l'objectif était d'améliorer la phonétisation d'un système de synthèse vocale de SMS en ce qui concerne trois types de problèmes : l'écriture rébus (chiffres et lettres utilisés pour leur valeur phonique), les abréviations sous forme de squelettes consonantiques et les agglutinations (déterminants ou pronoms collés graphiquement au mot qui suit). Notre approche se base sur l'analyse d'un corpus de SMS, à partir duquel nous avons extrait des listes de formes permettant de compléter les lexiques du système, et mis au point de nouvelles règles pour les grammaires internes. Les modifications effectuées apportent une amélioration substantielle du système, bien qu'il reste, évidemment, de nombreuses autres classes de problèmes à traiter.

This article presents a study whose goal is to improve the grapheme-to-phoneme component of an SMS-to-speech system. The three types of problems tackled in the study are: rebus writing (digits and letters used for their phonetic value), consonant skeleton abbreviations and agglutinations (determiner or pronouns linked with the next word). Our approach is based on the analysis of an SMS corpus, from which we extracted lists of forms to enhance the system's lexicons, and developed new grammatical rules for the internal grammars. Our modifications result in a substantial improvement of the system, although, of course, there remain many other categories of problems to address.

¹ Cette étude a été réalisé dans le cadre d'un contrat de recherche avec France Télécom R&D.

1 Introduction

La synthèse automatique de la parole à partir de texte consiste à transcrire un document écrit en son équivalent parlé. Les dispositifs actuels permettent de produire un signal acoustique de qualité jugée suffisante pour des applications nombreuses dans des domaines tels que l'aide aux personnes handicapées, le monitoring vocal, la communication homme/machine ou encore les services de télécommunication. Cependant, les progrès potentiels restent vastes tant en termes d'amélioration du naturel de la voix qu'en termes d'amélioration de la restitution du contenu sémantique. Un effort particulier doit être réalisé sur l'analyse linguistique pour parvenir à la vocalisation automatique d'un plus grand éventail de textes et en particulier celles des textes dits « mal formés » — il conviendrait sans doute de dire plutôt « non-standards » — comme ceux issus des SMS.

Cette communication rend compte de notre collaboration avec l'équipe du laboratoire Langues Naturelles (LN) au sein de France Télécom Recherche et Développement (FT R&D, pôle basé à Lannion, Côte d'Armor) sur le projet « *SMS2Voice* », dispositif de vocalisation de SMS développé par France Télécom R&D sous maîtrise d'ouvrage partagée Orange et FDF (Fixe et Distribution France). Le but de notre intervention était de tenter d'améliorer au mieux la couverture actuelle du système, en étudiant le fonctionnement du dispositif et en ciblant les traitements à effectuer. Les résultats de ce travail sont exposés dans ce papier.

2 Synthèse de la parole et SMS

Depuis 1965 et l'apparition sur le marché commercial d'un système de synthèse vocale développé par IBM, les programmes informatiques visant à synthétiser de la parole n'ont pas cessé de s'améliorer et de se multiplier (D'Alessandro, 2001). Toutefois, ces différents logiciels ne sont pas tous conçus pour analyser des textes « non-standards » comme le sont les messages envoyés par SMS.

Le SMS (acronyme de *Short Message Service*) est un service d'échange de messages écrits (limités à 160 caractères), proposé par tous les opérateurs de téléphonie mobile. La vocalisation automatique de SMS a de nombreuses applications. Elle intéresse tout d'abord les différents opérateurs de téléphonie mobile pour la réception de SMS sur poste fixe. La synthèse vocale à partir de SMS pourrait également rendre ceux-ci accessibles aux aveugles et malvoyants. Enfin, la technologie aurait un intérêt non négligeable pour tous les métiers où les mains sont occupées (par exemple, les chauffeurs routiers) et pour lesquels il serait plus pratique de pouvoir écouter ses messages plutôt que de les lire.

Lorsqu'on s'intéresse à ce type de messages, on se rend rapidement compte que ceux-ci présentent de nombreuses particularités linguistiques problématiques pour la vocalisation (Anis, 2001, 2002 ; Guimer de Neef & Véronis, 2004). Il suffit pour s'en convaincre d'écouter la sortie produite par un synthétiseur sur des messages tels que :

dsl ma bel! ms c dernié tps javé pa tro le tps
(*Désolé ma belle ! Mais ces derniers temps j'avais pas trop le temps.*)

slt jsp ke tt va bil
(*Salut, j'espère que tout va bien.*)

Les procédés d'écriture employés dans les SMS sont divers et nombreux (nous renvoyons le lecteur à Anis, 2002, pour une étude détaillée). Dans cette communication nous exposons uniquement trois phénomènes spécifiques de l'écriture SMS, que nous avons étudiés en collaboration avec France Télécom : l'écriture *rébus* (chiffres et lettres), l'écriture par *squelettes consonantiques*, et le procédé d'*agglutination*, dont nous pensons qu'ils étaient à l'origine de nombreuses erreurs de traitement par leur analyseur linguistique (que nous présentons plus bas). Nous décrivons brièvement ces procédés ci-dessous.

2.1 L'écriture rébus

Nous entendons par « rébus » le procédé d'écriture par lequel certaines séquences de lettres sont remplacées par un arrangement de chiffres et/ou de lettres correspondant au même phonème que la séquence en question. Exemples² :

2m1 = *demain* [rébus chiffre]

Kfé = *café* [rébus lettre]

2.2 Squelettes consonantiques

Nous considérons comme squelettes consonantiques les mots dont les voyelles ont été supprimées, réduisant ainsi la forme à une succession des consonnes principales du mot. Comme le souligne justement (Anis, 2002), nous savons depuis longtemps grâce à la théorie de l'information, que les consonnes ont une valeur informative plus forte que les voyelles. Le mot français écrit est fortement charpenté autour des consonnes, dont certaines n'ont pas de contrepartie phonique. Exemples :

slt = *salut*

prt = *pourtant* (notons que le « n » du son « an » n'a pas été conservé alors que le « t » muet final l'a été)

tjs = *toujours* (de la même manière ici le « r » pourtant prononcé n'est pas conservé alors que le « s » muet l'est)

2.3 Les agglutinations

Nous appelons enfin « agglutination » la formation d'un mot par la réunion de deux ou plusieurs unités lexicales (*jpouré* = « je pourrai »). Il est à noter que nous nous avons traité uniquement le cas des agglutinations binaires (combinaison de deux unités). Le tableau 1 donne un échantillon d'exemples pour les séquences d'agglutinations avec clitiques.

² Les exemples sont extraits du corpus sur lequel nous avons travaillé et que nous présentons plus loin.

Forme agglutinée	Exemples de patrons observés	Exemples de formes
« JE »	j + pronom j + verbe	jte (= je te) jsui (= je suis)
« QUE » (sous la forme « KE »)	k + <i>article défini</i> k + <i>pronom démonstratif</i> k + <i>pronom personnel</i>	kle (= que le) kce (= que ce) ktu (= que tu)
« CE / SE »	s + pronom s + verbe	ske (= ce que) svoir (= se voir)
« ME »	m + pronom m + verbe	mle (= me le) mparl (= me parle)

Tableau 1 : Exemples d'agglutinations

3 Méthodologie

Nous avons tout d'abord fait un relevé systématique des erreurs de traitement et de vocalisation produites par SMS2Voice. Le dispositif actuel était opérationnel sur de nombreux points, mais son fonctionnement pouvait être amélioré pour certains phénomènes étant encore incorrectement ou nullement traités, tels que :

gcrai (j'essaierai) vocalisé tel quel.
attd (attend) vocalisé tel quel.
Deboulot (de boulot) traduit « *déboule* ».

A la suite des observations, nous nous sommes rendus compte que les systèmes de correction et les lexiques préalablement en place présentaient certaines limites. Il semblait donc important de poursuivre cette étude avec un travail sur corpus afin d'approfondir la connaissance des phénomènes SMS (plus particulièrement les procédés choisis pour cette étude) et améliorer le système.

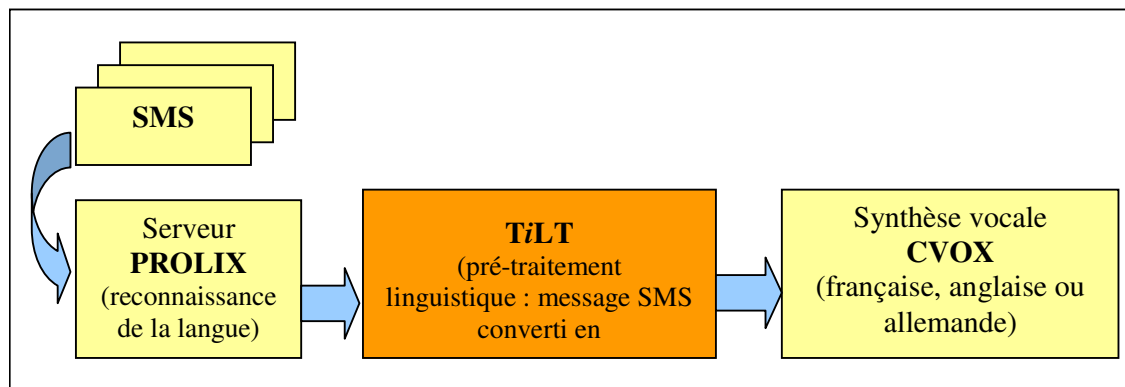


Figure 1 : Architecture générale du système SMS2Voice

Le premier prérequis important à cette étude était de disposer d'un corpus à partir duquel il était possible de faire un certain nombre d'observations et de traitements. Le corpus de SMS utilisé dans cette étude a été réalisé par des étudiants de l'Université de Provence de 2000 à

2004 dans le cadre de travaux pratiques et de mémoires. Celui-ci contient 13 400 messages représentant près de 156 620 mots. Nous avons soumis ce corpus au dispositif *SMS2Voice*. La figure 1 représente globalement la structure de l'application. Notre travail a porté sur la brique logicielle *TiLT* (TLiT → TiLT, **T**raitement **L**inguistique de **T**extes ; cf. Guimier de Neef *et al.* 2002), déjà en utilisation chez France Télécom pour de nombreux projets (indexation et filtrage linguistique, résumé automatique, classement thématique, etc.), et qui constitue une véritable « boîte à outils » pour le Traitement Automatique des Langues. Un module permet de segmenter un texte d'entrée en mots ; ceux-ci sont ensuite soumis à une analyse lexicale, morphologique, et éventuellement des corrections. Ces diverses informations sont ensuite traduites afin d'affecter une catégorie grammaticale à chaque type de mot, et pour pouvoir procéder à une analyse syntaxique et sémantique des données ainsi traitées. Pour qu'il puisse fonctionner il lui faut donc :

- Des **données de segmentation** (pour le découpage d'un texte en phrases, mots, numéros, signes de ponctuation etc.).
- Des **lexiques** (avec informations morpho-flexionnelles pour l'association de chaque mot à ces différentes analyses hors contexte)
- Des **stratégies de corrections** (pour la correction de formes erronées)
- Une **traduction de traits** (qui fait la correspondance entre les traits lexicaux et les étiquettes grammaticales)
- Des **grammaires** (pour la désambiguïsation lexicale par exploration du contexte)

Ce système est initialement prévu pour analyser des textes au format « standard » et nous avons participé à son adaptation aux traitements de textes non-standards.

Pour commencer, nous avons dégagé des listes d'occurrences représentant les mots les plus fréquents du corpus selon divers critères d'identification, afin d'étudier notamment si les formes les plus courantes étaient les plus problématiques ou non. Pour ce faire, nous avons comparé les listes de mots extraites du corpus avec un lexique de français standard de référence, le lexique MULTEXT (Ide & Véronis, 1994). Le but de notre approche était de voir pour chaque phénomène donné quelles sont les formes que le système est déjà en mesure de traiter, et d'isoler ainsi les occurrences problématiques pour leur appliquer le traitement adéquat en fonction de leurs particularités (longueur, complexité, etc.).

4 Traitement du corpus

4.1 Extraction des occurrences

La première étape pour étudier chacun des procédés nécessite d'extraire le plus précisément possible du corpus la liste des occurrences correspondant aux critères d'identification du phénomène, par l'intermédiaire de scripts. Pour ce faire, nous avons mis en place des traitements qui permettent, à partir de la liste des formes inconnues en français standard, d'extraire le plus précisément possible les formes voulues. L'opération suivante consiste à éliminer manuellement les formes qui ont été extraites par le script sur une base purement formelle, mais ne correspondent pas au phénomène recherché. En effet, la majorité des phénomènes est difficile à filtrer et à extraire directement, et dans un premier temps nous ne pouvons nous baser que sur des indices pour les identifier.

Par exemple pour les agglutinations, nous nous basions sur le fait que les formes recherchées commencent par une succession de deux consonnes (dont la première est, par exemple, une forme clitique élidée), suivies de n'importe quel caractère. Or, bien que cette approche nous permette d'extraire des formes telles que « jsui », « ktu » ... qui répondent effectivement au patron recherché, elle nous amène aussi à extraire les occurrences « tro », « spectacl », « jamé » (= jamais)... qui correspondent également au patron mais qui ne sont pas des formes agglutinées.

4.2 Étude quantitative

A la suite de l'étape d'extraction des occurrences, nous avons donc été contraint de procéder à une phase de tri manuel des formes extraites pour ne garder que celles correspondant au patron voulu. Le tableau 2 donne, pour chacun des procédés étudiés, les têtes de listes des occurrences (ainsi que leur fréquence d'apparition) lors de la première extraction, puis la liste des occurrences restantes à traiter après filtrage manuel.

TÊTES DE LISTE POUR LES REBUS CHIFFRES			
1^{ère} extraction de formes		Formes restantes après filtrage manuel	
<i>Fréquence</i>	<i>Occurrence</i>	<i>Fréquence</i>	<i>Occurrence</i>
63	2m1	63	2m1
62	bi1	62	bi1
35	2min	35	2min
25	2main	25	2main
17	2pui	17	2pui
13	b1	13	b1
13	1er	12	vi1
12	vi1	10	qq1
12	2m	7	dem1
10	qq1	6	pl1

TÊTES DE LISTE POUR LES REBUS LETTRES			
1^{ère} extraction de formes		Formes restantes après filtrage manuel	
<i>Fréquence</i>	<i>Occurrence</i>	<i>Fréquence</i>	<i>Occurrence</i>
29	reCu	9	paC
26	MonNuméro	4	pE
9	paC	3	mR
4	pE	2	vE
4	gpVeuilz	2	trouV
4	faCon	2	jaV
3	mR	2	jV
2	vE	2	danC
2	trouV	2	creV
2	reCus	2	cT

TÊTES DE LISTE POUR LES SQUELETTES CONSONANTIQUES			
1 ^{ère} extraction de formes		Formes restantes après filtrage manuel	
Fréquence	Occurrence	Fréquence	Occurrence
209	sms	167	slt
167	slt	145	svp
145	svp	120	stp
120	stp	52	qqch
52	qqch	48	lgtps
48	lgtps	39	rdv
39	rdv	37	bcp
37	bcp	35	jrs
35	jrs	33	msg
33	msg	29	tps

TÊTES DE LISTE POUR LES AGGLUTINATIONS			
1 ^{ère} extraction de formes		Formes restantes après filtrage manuel	
Fréquence	Occurrence	Fréquence	Occurrence
209	sms	55	jsui
167	slt	42	jte
145	svp	23	jt
144	tro	20	jp
132	ds	14	jme
120	stp	12	jm
97	st	11	ktu
64	ms	10	lpe
63	2m1	9	jvai
60	ns	9	jtapel

Tableau 2 : Têtes de liste des occurrences pour les différents procédés

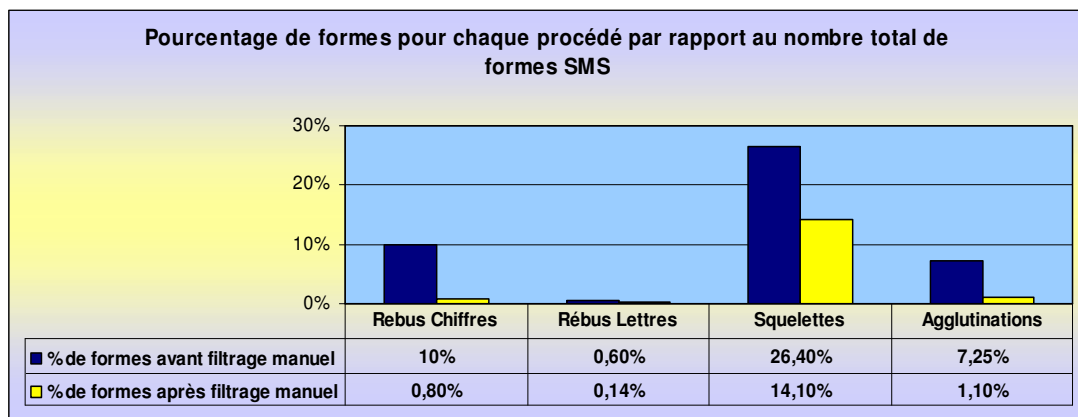


Figure 2 : Filtrage manuel

La figure 2 donne le récapitulatif statistique de nos observations pour chacun des phénomènes étudiés, une fois le tri manuel effectué. On note qu'une grande partie des formes

correspondant aux patrons d'extraction ne correspond pas aux procédés recherchés, ce qui justifie une analyse linguistique manuelle. Un système automatique basé sur de tels patron aurait des performances très faibles.

5 Amélioration du système

Dès que les formes correspondant au procédé sont identifiées, il est nécessaire de vérifier si elles sont connues et correctement traitées par le système ou non, afin d'améliorer un certain nombre de ressources du système de manière à le rendre plus opérationnel. Pour cela nous avons recherché (à l'aide de scripts en langage Perl) les listes d'occurrences extraites et non traitées par le système. Lorsque ces formes inconnues sont isolées, une stratégie de traitement doit être adoptée. Deux possibilités sont offertes (la décision dépend généralement du phénomène traité) :

- soit encoder les formes dans le fichier des formes figées du domaine SMS ;
- soit modifier un certain nombre de règles et de données (phonétiques, lexicales, etc.)

Pour le cas des squelettes consonantiques par exemple, les formes inconnues du système (*slt*, *bcp*, etc.) ont été simplement encodées dans le fichier des formes figées, mais des règles ont été créées pour les rébus et agglutinations

Concernant le procédé d'écriture rébus (chiffres et/ou lettres), nous avons ajouté de nouvelles règles phonétiques. Par exemple, une règle de type $1 \rightarrow j\hat{a}$ donne la possibilité pour le chiffre « 1 » de correspondre à la séquence « ien » (exemple : « rev1 » = reviens ; « ch1 » = chien ; etc.). Ce type de modification permet donc de générer un nouveau phonétiseur et d'améliorer ainsi la sortie produite lors de l'accès au lexique (Tableau 3).

Avant modification du phonétiseur	Après modification du phonétiseur
<pre>> v1 Mot analysé : v1 Nombre de solutions : 10 v1 vain vainc vaincs vains vin vingt vins vint vînt</pre>	<pre>> v1 Mot analysé : v1 Nombre de solutions : 12 v1 vain vainc vaincs vains viens vient vin vingt vins vint vînt</pre>

Tableau 3 : Exemple d'accès au lexique

En ce qui concerne les agglutinations, notre intervention a porté sur différents niveaux. En effet, les formes agglutinées de moins de 5 caractères (seuil à partir duquel nous supposons que le traitement par règles sortirait un nombre trop important de formes erronées) ont été directement encodées dans le fichier de formes figées. En revanche, pour des formes dont la

longueur est supérieure à ce seuil nous avons modifié et/ou ajouté des règles de grammaire adaptées. Exemples :

((LETTRE_ELIDEE ADV_NEG_NE) (VERB))
(pour le cas d'agglutination entre « **ne** » et un **verbe**, ex. : *nviendrai*)

((LETTRE_ELIDEE PREP) (VERB INF))
(pour le cas d'agglutination entre **préposition** et **verbe à l'infinitif**, ex. : *dpartir*)

Il convient de préciser qu'une fois les traitements effectués, il est nécessaire de recompiler un certain nombre de données (en fonction des fichiers modifiés). Nous avons ensuite effectué différents tests de «non-régression» pour vérifier que les modifications opérées ne génèrent pas de dysfonctionnements importants du système. Ces tests consistent à appliquer TiLT sur une série de fichiers spécifiques (ex : messages avec systématiquement le chiffre 2, la lettre d ou encore la forme interrogative « *eske* » [est-ce que]). Les fichiers de sortie permettent notamment de voir la différence entre les résultats de l'ancien test et ceux de celui venant d'être effectué. Cette phase est particulièrement importante car elle permet de s'assurer qu'il n'y a pas eu de problème majeur avant de passer à d'autres modifications sur les données, et elle permet de juger de la pertinence des améliorations apportées.

6 Conclusions / perspectives

Le travail réalisé a permis d'améliorer de façon significative la performance du système *SMS2Voice*. Une évaluation quantitative détaillée pourrait être réalisée, mais sa mise en œuvre demanderait un effort important qui dépassait le cadre qui nous était imparti. De plus, une telle évaluation serait plus pertinente à entreprendre lorsque d'autres phénomènes auront pu être traités. Par une analyse qualitative systématique, nous avons en effet pu mettre en évidence de nombreux points qui méritent eux aussi amélioration. Au-delà des interventions que nous avons décrites, nous avons pu proposer différentes grilles d'analyse et des typologies (morpho-lexicale, morpho-syntaxique, etc.) qui permettront de faciliter la suite de ce travail.

Quelques points restent à améliorer pour les procédés que nous avons traités. Concernant le phénomène d'agglutination notamment, diverses règles grammaticales peuvent être encore ajoutées (par exemple, les observations menées sur les agglutinations binaires méritent d'être étendues au cas des agglutinations ternaires (ex : *jtsouhaite* [je te souhaite])). Il reste également un nombre important de formes et de procédés listés par Anis (2003) que nous n'avons pas abordés et dont le traitement s'annonce extrêmement délicat. Par exemple, une forme telle que « *ldpdte* » (indépendante) fait appel à trois procédés cumulés :

1. Rébus avec chiffre : « 1 » (in-)
2. Rébus avec lettre : « d » (-dé-)
3. Squelette consonantique : « *pdte* » (-pendante)

Ce type de combinaison pose des problèmes de phonétisation difficilement résolubles dans l'état actuel des connaissances. En effet, pour traiter une telle occurrence le système TiLT devrait avoir recours à diverses stratégies de correction. La difficulté vient de l'architecture modulaire du système (qui est d'ailleurs commune à la plupart des systèmes de TAL actuels) : chaque module de correction intégré à l'analyseur linguistique intervient de façon indépendante, séparément des autres modules. Il n'y a pas de traitement global pour

l'application des corrections, et l'application systématique de toutes les combinaisons de modules produit une explosion combinatoire extrêmement coûteuse et néfaste *in fine* à la précision du système.

Il demeure donc de nombreux aspects intéressants pour des études futures. L'écriture SMS n'est ni normée, ni stable ; une vocalisation de qualité nécessite donc une explicitation de tout le message textuel quels que soient les procédés d'écriture employés. Il est donc certainement nécessaire et important de poursuivre l'approfondissement des travaux de linguistique et de traitement automatique menés sur les données issues de SMS, ainsi que d'autres formes d'écrit non standards qui utilisent des procédés analogues (chats et e-mails, cf. Torzec, 2001).

Remerciements

Je tiens à remercier Emilie Guimier de Neef qui a encadré ce stage au sein de France Télécom, ainsi qu'à toute l'équipe du laboratoire Langues Naturelles qui m'a accueilli. J'adresse également tous mes remerciements à Jean Véronis qui a dirigé ce travail. Je leur suis reconnaissant pour leurs nombreux commentaires sur cet article.

Références

- Anis, J. (2001). (Dir.), (2001), « *Parlez-vous texto ?* », Paris : Le Cherche Midi Éditeur.
- Anis, J. (2002), Communication électronique scripturale et formes langagières : chats et SMS. *Actes des Quatrièmes Rencontres Réseaux Humains / Réseaux Technologiques*, Université de Poitiers. [<http://oav.univ-poitiers.fr/rhrt/2002/actes%202002/jacques%20anis.htm>]
- D'Alessandro. C. (2001). 33 ans de synthèse de la parole à partir de texte : une promenade sonore (1968-2001), *Traitement automatique de la parole*, Vol. 42(1), pp. 297-321.
- Guimier De Neef E., Boualem M., Chardenon C., Filoche P., Vinesse J. (2002), Natural Language processing software tools and linguistic data developed by France telecom R&D, *CDAC Conference*, India.
- Guimier de Neef, E. & Véronis, J. (2004). 1 pw1 sr la kestion ;-). *Journée d'Étude de l'ATALA "Le traitement automatique des nouvelles formes de communication écrite (e-mails, forums, chats, SMS, etc.)"*, Paris.
- Ide, N., & Véronis, J. (1994). MULTEXT (Multilingual Tools and Corpora), *14th International Conference on Computational Linguistics, COLING'94*. Kyoto. 588-592.
- Torzec, N., Moundenc, T., Emerard, F. (2001). Prétraitement et analyse linguistique dans le système de synthèse TTS CVOX : Application à la vocalisation automatique d'e-mails, *Traitement automatique de la parole*. Vol. 42(1), pp. 17-46.