



HAL
open science

Développement et évaluation d'un système de synthèse pour l'espagnol vénézuélien : projet et état d'avancement

Elsa Mora, Daniel J. Hirst, Christian Cavé

► To cite this version:

Elsa Mora, Daniel J. Hirst, Christian Cavé. Développement et évaluation d'un système de synthèse pour l'espagnol vénézuélien : projet et état d'avancement. Travaux interdisciplinaires du Laboratoire Parole et Langage, 2000, 19, pp.91-98. hal-00285403

HAL Id: hal-00285403

<https://hal.science/hal-00285403>

Submitted on 5 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DÉVELOPPEMENT ET ÉVALUATION D'UN SYSTÈME DE SYNTHÈSE POUR L'ESPAGNOL VÉNÉZUÉLIEN : PROJET ET ÉTAT D'AVANCEMENT¹

Elsa Mora², Daniel Hirst, Christian Cavé

Résumé

Nous développons actuellement un système de synthèse de parole à partir du texte, basé sur la concaténation de diphtongues, pour l'espagnol vénézuélien. Nous présentons les grandes lignes du projet initial et l'état d'avancement des travaux. La base complète de diphtongues a été enregistrée, pour le moment avec un seul locuteur masculin. La segmentation et l'étiquetage des diphtongues sont en cours. À côté de la réalisation du système de synthèse lui-même, le programme de travail prévoit la mise au point de plusieurs corpus et de procédures permettant d'évaluer l'intelligibilité et la qualité de la parole produite.

Abstract

This paper reports on a joint project between the Laboratoire Parole et Langage, Aix-en-Provence and the Department of Linguistics of Mérida (Venezuela) concerning the development of a Text to Speech synthesis System for Venezuelan Spanish based on diphone concatenation. The project is financed by Ecos-Nord (France) and Conicit (Venezuela). A complete diphone set for Venezuelan Spanish has been recorded and is at present in the process of being segmented and labeled. Besides the production of a diphone database, the project involves the recording of a number of different speech corpora and the development of procedures for the evaluation of the intelligibility and the quality of the synthetic speech.

Mots-clés : synthèse à partir du texte, espagnol vénézuélien, diphtongue.

Keywords : Text to Speech System, Venezuelan Spanish, diphone.

Introduction

Nous présentons ici les points principaux d'un projet de recherche en cours et son état d'avancement. Ce projet résulte de la réponse conjointe du Département de Linguistique de Mérida (Venezuela) et du Laboratoire Parole et Langage d'Aix-en-Provence à un appel à collaboration bilatérale d'Ecos-Nord. Il est financé à part égale par Ecos-Nord (France) et par Conicit (Venezuela) pour une durée de 4 ans.

¹ Programme soutenu par Ecos-Nord (Action V99H01), France, Conicit et Fundayacucho, Venezuela.

² Département de Linguistique, Université des Andes, Mérida, Venezuela.

Ce projet concerne la synthèse de la parole dont le champ d'application ne cesse de s'étendre. Il comprend deux parties : l'une porte sur le développement d'un système de synthèse de parole pour l'espagnol vénézuélien ; l'autre sur des procédures et des outils permettant d'évaluer l'intelligibilité et la qualité de la parole de synthèse obtenue.

L'intérêt de ce projet pour le laboratoire Parole et Langage est multiple. Il permettra de parfaire son expérience dans le domaine de la synthèse de la parole multilingue à partir du texte, domaine considéré comme prioritaire pour l'équipe Prosodie et Représentation Formelle du Langage. Il permettra également d'étendre l'application de son modèle formel de la prosodie à l'espagnol vénézuélien, après avoir été appliqué à l'anglais et au français. La deuxième partie du projet permettra en outre de vérifier l'applicabilité des tests de perception mis au point pour le français sur une autre langue romane.

I. Contexte du projet

La synthèse à partir du texte est actuellement en train de faire des progrès considérables dans de nombreuses langues grâce au projet MBROLA mis en place par Thierry Dutoit de l'Université Polytechnique de Mons en Belgique (Dutoit et al., 1996; Dutoit, 1997). Ce système de synthèse vocale permet de produire un énoncé naturel par concaténation de petits segments de parole préenregistrés en spécifiant simplement la nature des unités et leurs caractéristiques prosodiques (durée et contour mélodique). Il est possible de faire la synthèse à partir du texte d'un énoncé quelconque d'une langue donnée en choisissant comme segments préenregistrés les diphtongues qui sont les transitions entre les parties stables d'une séquence de deux phonèmes. Par exemple, l'énoncé « il est parti » /ileparti/ sera produit en concaténant les diphtongues suivants : #i; il; le; ep; pa; ar; rt; ti; i#.

La mise à disposition gratuite de cet outil auprès de la communauté internationale a encouragé de nombreux laboratoires dans plusieurs pays à constituer des bases de diphtongues pour procéder au développement d'un système de synthèse à partir du texte et à effectuer des recherches sur les caractéristiques prosodiques des langues concernées de façon à améliorer la qualité et le naturel de la synthèse obtenue.

Après constitution de la base de diphtongues, le synthétiseur est mis gratuitement à la disposition des utilisateurs potentiels pour toute application non commerciale et non militaire. Les bases de diphtongues d'origine restent la propriété entière des laboratoires qui les ont constitués. Les bases de diphtongues compatibles avec MBROLA sont actuellement disponibles pour 20 voix dans 13 langues différentes dont l'espagnol européen et l'espagnol mexicain (Voir <http://tcts.fpms.ac.be/synthesis/mbrola.html>). Il n'existe pas encore de base de diphtongues ni de synthétiseur pour l'espagnol vénézuélien.

Réaliser des systèmes de synthèse pour les différentes variétés d'une même langue se justifie pour deux raisons. La première est justement de tenir compte des variantes dialectales. Ainsi, l'espagnol d'Amérique du Sud se différencie du péninsulaire par un certain nombre de caractéristiques phonétiques, phonologiques et prosodiques. De même, un système de synthèse pour le français ne «parlera» pas de la même façon selon qu'il est français, suisse ou belge. La deuxième raison est liée à l'affirmation d'une identité culturelle dont la langue, dans ses variantes dialectales, est un élément important. C'est ainsi, par exemple, qu'il existe un projet de base de diphtonges pour le français méridional (Chaumette, Clairet, Coquillon, 1997) (cf. <http://www.lpl.univ-aix.fr/projects/index.html>).

L'équipe d'Aix a développé un système de synthèse à partir du texte (Di Cristo, Di Cristo, Véronis, 1997), comportant des règles prosodiques s'appuyant sur des études théoriques de la représentation multilingue de la prosodie (Hirst, Di Cristo, 1998; Mora, 1996). L'adaptation de ce système à d'autres langues ou à d'autres variétés d'une langue est en cours de développement.

1.1. Mesure de l'intelligibilité de la parole

La mesure de l'intelligibilité de la parole nécessite de disposer de corpus spécifiques. Les résultats les plus précis sont obtenus avec un corpus permettant de mesurer l'intelligibilité segmentale au niveau des traits. Il est donc nécessaire de mettre au point, pour l'espagnol vénézuélien, un test de paires minimales aussi appelé test de rime ou DRT (Diagnostic Rhyme Test). Ce type de test existe actuellement, à notre connaissance, pour l'anglais, le néerlandais et pour le français ainsi que pour l'espagnol péninsulaire. Le principe de ce test est d'opposer des paires de mots différant à l'initiale par un seul trait distinctif (par exemple pile/bile -trait de voisement- ou tâche/sache -trait interrompu). L'avantage de ce type de test est qu'il fournit d'une part une mesure globale de l'intelligibilité et d'autre part une mesure diagnostique informant sur le type d'erreur.

Le test de paires minimales dépasse largement le cadre strict de la synthèse de parole puisqu'il permet aussi bien de mesurer la parole codée (utilisation de vocodeur), l'efficacité d'un système de transmission ou les performances d'une prothèse auditive ou encore les erreurs de décodage propre à un auditeur particulier.

1.2. Évaluation de la qualité de la parole synthétique

Les progrès réalisés dans les systèmes de synthèses ont amené à dépasser la seule évaluation de l'intelligibilité pour s'intéresser à la qualité de la parole à laquelle la prosodie contribue fortement. L'évaluation de la prosodie est un thème majeur de ces dernières années faisant l'objet de plusieurs actions internationales (AUPELF, COST 258) bien qu'aucune méthode ne se soit imposée.

2. Méthodologie

La première condition pour réaliser la synthèse d'une langue quelconque en utilisant MBROLA est de créer une base de diphtonges de la langue considérée. Cela nécessite plusieurs opérations.

2.1. Corpus

Le préalable à la création de ce corpus est de faire l'inventaire exhaustif des diphtonges, ce qui contribue à faire progresser la description de la langue. Le nombre de diphtonges est de l'ordre de grandeur du carré du nombre de phonèmes de la langue.

2.2. Enregistrement

Le corpus sera enregistré par au moins un locuteur de langue maternelle espagnol vénézuélien.

2.3. Traitement

Cette étape consiste à numériser et à étiqueter tous les diphtonges de la langue. Cet ensemble constitue la base de diphtonges proprement dite. Elle est ensuite compilée de façon à pouvoir être utilisée directement par le logiciel MBROLA.

3. Déroulement des opérations

3.1. La création d'une base de diphtonges est une opération relativement longue et complexe. Elle est habituellement réalisée en plusieurs étapes. Nous proposons la création d'un synthétiseur pour l'espagnol vénézuélien avec les étapes suivantes :

Étape 1

Réalisation d'une "petite" base de quelques dizaines de diphtonges. L'intérêt principal de cette étape est de pouvoir synthétiser quelques phrases-tests afin d'évaluer la qualité de la parole obtenue et donc de faire éventuellement des mises au point nécessaires avant d'enregistrer la totalité du corpus.

Étape 2

Enregistrement du corpus complet, vérification de la qualité des enregistrements.

Étape 3

Compilation de la base de diphtonges ; intégration dans MBROLA. Cette étape est réalisée par l'équipe de Mons qui ensuite détruit la copie de la base qui leur a été transmise. On dispose alors d'un synthétiseur vocal pour l'espagnol vénézuélien.

Un synthétiseur vocal est, bien entendu, un élément central pour la constitution d'un système complet de synthèse à partir du texte écrit. Il présente aussi un intérêt pour la recherche fondamentale dans la mesure où il permet de générer des corpus et donne des possibilités de modification du signal de parole pour des expériences sur la perception de la parole. En outre un synthétiseur de ce type peut être utilisé pour l'alignement automatique de transcriptions phonématiques avec un signal de parole, ce qui en fait un outil précieux pour les recherches sur la prosodie.

Étape 4

Cette étape consistera à compléter le synthétiseur vocal par des règles de génération automatique de prosodie en intégrant des modules inspirés du système SYNTAIX développée à Aix pour le français (Di Cristo et al., 1997) et s'appuyant sur des recherches sur la représentation multilingue de la prosodie effectuées par l'équipe d'Aix depuis de nombreuses années (Hirst, Di Cristo, 1998; Hirst et al., 2000). À cette fin, le laboratoire mettra à la disposition des partenaires vénézuéliens les outils de modélisation prosodique et de traitement du signal de parole qui ont été développés. Il sera nécessaire de constituer un corpus de référence pour le développement des règles spécifiques de la prosodie de l'espagnol vénézuélien (Mora, 1996).

À côté du développement d'un système de synthèse, il est indispensable de développer des outils permettant de mesurer l'intelligibilité et la qualité de la parole. Le laboratoire Parole et Langage a mis au point une station de test automatisée pour la présentation des stimulus et le recueil des résultats (Cavé et al, 1995, 1997). Il reste à mettre au point un corpus spécifique en espagnol vénézuélien.

3.2. Évaluation de la prosodie

Les méthodes indirectes et en particulier la méthode de vérification de phrases («sentence verification task»; SVT) semblent très prometteuses. Le principe consiste à présenter au sujet de courtes phrases auxquelles on peut répondre par OUI-NON ou par VRAI-FAUX et à mesurer le temps de réponse. L'hypothèse est que le temps de réponse est dépendant de la qualité de la parole dans la mesure où une parole de moins bonne qualité oblige à un effort perceptif ou cognitif allongeant le temps de traitement. Cette technique ainsi que d'autres sont en cours d'expérimentation à Aix dans le cadre d'une action spécifique du projet AUPELF. Les techniques adoptées pour le français seront testées sur l'espagnol vénézuélien.

4. Calendrier des opérations

•ière année : constitution de l'inventaire des diphtongues. Réalisation d'une maquette permettant la synthèse et l'évaluation informelle de quelques phrases choisies au préalable.

•2ème année : constitution de la base de diphtongues complète. Mise au point d'un corpus de paires minimales en espagnol vénézuélien. Mise au point, validation et étalonnage des tests d'intelligibilité.

•3ème année : réalisation du système de synthèse vocale. Constitution d'un corpus de référence pour étude de la prosodie. Étiquetage automatique du corpus au moyen du synthétiseur. Mise au point d'un corpus spécifique, en espagnol vénézuélien, pour la procédure SVT.

•4ème année : Réalisation d'un système de synthèse par règles à partir du texte avec module prosodique spécifique à l'espagnol vénézuélien. Validation et étalonnage des tests d'évaluation.

5. Résultats escomptés à terme

- Une base de diphtongues pour l'espagnol vénézuélien.
- Un système de synthèse vocale permettant la création de stimulus artificiels pour des tests de perception.
- Un outil d'alignement automatique pour l'espagnol vénézuélien.
- Un système de synthèse par règles à partir du texte pour l'espagnol vénézuélien avec des règles prosodiques spécifiques.
- Corpus et procédure pour la mesure de l'intelligibilité et de la qualité de la parole en espagnol vénézuélien.

6. État d'avancement

6.1. Prémaquette

Il est prudent lorsqu'on se lance dans la création d'une base de diphtongues de réaliser d'abord une prémaquette à partir d'une mini-base de diphtongues. Cela permet d'optimiser les procédures d'enregistrement et de segmentation de façon à avoir une base complète de bonne qualité. Pour les premiers essais, nous avons choisi comme phrase à synthétiser un proverbe populaire « no dejes para mañana lo que puedas hacer hoy »³ et nous avons établi la liste de diphtongues nécessaires à la génération de cette phrase. Pour chaque diphtongue, on a trouvé un mot courant ou une courte phrase contenant ce diphtongue. Le corpus ainsi obtenu a été inséré dans une phrase porteuse « él dijo ... » et a été enregistré par deux locuteurs et une locutrice. On a ensuite segmenté et étiqueté les diphtongues correspondants et réalisé la synthèse par concaténation de la phrase sélectionnée. La mini-base de diphtongues est ensuite compilée par l'équipe de Mons pour obtenir une sortie vocale en synthèse MBROLA, ce qui permet d'évaluer la qualité de cet échantillon de diphtongues.

³ « Ne laisse pas pour demain ce que tu peux faire aujourd'hui. »

6.2. Base de diphtongues

On a établi la liste complète des diphtongues de l'espagnol vénézuélien en considérant tous les phonèmes et les variantes les plus courantes des différents phonèmes. Pour les consonnes, on a pris en compte les variantes occlusives et fricatives des occlusives bilabiales, dentales et vélares ; le phonème fricatif alvéolaire sourd en position initiale et son allophone fricatif glottal qui se réalise généralement en position finale de syllabe. À part ce cas, on n'a pas considéré les différences possibles entre initiale et finale de syllabes. Pour les voyelles, on a pris en compte la réalisation accentuée et inaccentuée ainsi que la réalisation en diphtongue. On aboutit ainsi à une liste de 30 réalisations phoniques (voir ci-dessous) qui sert de base pour l'établissement de la liste des diphtongues qui seront présents dans la base complète. Compte tenu de la prise en compte de réalisations allophoniques et de suites de phonèmes impossibles, on a obtenu une liste de 674 diphtongues. L'enregistrement de la base complète vient d'être réalisé (décembre 2000).

Liste des réalisations phoniques retenues pour la base de diphtongues
(codage SAMPA, sauf cas particuliers)

Consonnes

p, b, t, d, k, g, tS, f, B, D, s (position initiale de syllabe), s2 (position finale de syllabe), h (cette réalisation correspond à la prononciation des graphies j, ge et gi ainsi qu'à la prononciation du phonème /s/ en fin de syllabe. Il n'existe pas de code SAMPA pour cette réalisation), G, m, n, J, l, L (l'espagnol vénézuélien a une seule réalisation pour les graphies «y» et «ll». Il s'agit d'une fricative palatale sonore pour laquelle il n'y a pas de code SAMPA), r, rr, j, w

Voyelles

a, a*, e, e*, i, i*, o, o*, u, u* (* indique une voyelle accentuée)

6.3. Corpus de paires minimales en espagnol vénézuélien

De façon à pouvoir évaluer l'intelligibilité de la parole en espagnol vénézuélien et les performances d'un synthétiseur au niveau segmental, il est indispensable de disposer d'un test de paires minimales. Il existe déjà un test de paires minimales pour l'espagnol péninsulaire. Nous procédons actuellement à l'adaptation de ce test à l'espagnol vénézuélien.

Bibliographie

- CAVÉ, C.; TESTON, B.; GHIO, A.; DI CRISTO, P. (1995). Une station de test automatisée pour la mesure de l'intelligibilité et de la qualité de la parole. *Travaux de l'Institut de Phonétique d'Aix*, 16, p. 123-140.
- CAVÉ, C.; TESTON, B.; GHIO, A.; DI CRISTO, P. (1997). A computer-driven system for assessing speech quality and intelligibility. *Acta Acustica*, 84, 1, p. 157-161.
- CHAUMETTE, C.; CLAIRET, S.; COQUILLON, A. (1997). *Réalisation d'une maquette de base de dipbones en français méridional : application à la synthèse vocale*. Mémoire annexe, DEA Langage et Parole (direction DJ. Hirst)
- DI CRISTO, A.; DI CRISTO, P.; VÉRONIS, J. (1997). A metrical model of rhythm and intonation for French text-to-speech synthesis. *Proc. ESCA Workshop on Intonation*, Athens, sept. 1997, p. 83-86.
- DUTOIT, T. (1997). *An Introduction to Text-to-Speech Synthesis*, Dordrecht: Kluwer.
- DUTOIT, T.; PAGEL, V.; PIERRET, N.; BATAILLE, F.; VAN DER VRECKEN, O. (1996). The MBROLA project. Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. *Proceedings ICSLP '96* (Philadelphia) 3, p. 1393-1396.
- HIRST, DJ.; DI CRISTO, A. (eds) (1998). *Intonation System. A Survey of Twenty Languages*. Cambridge : Cambridge University Press.
- HIRST, DJ.; DI CRISTO, A.; ESPESSER, R. (2000). Levels of representation and levels of analysis for the description of intonation systems. In M. Horne (ed.) *Prosody: Theory and Experiment*, Dordrecht : Kluwer Academic Publishers.
- MORA, E. (1996). *Caractérisation prosodique de la variation dialectale de l'Espagnol parlé au Vénézuéla*. déc. 1996, Thèse de Doctorat, Université de Provence.