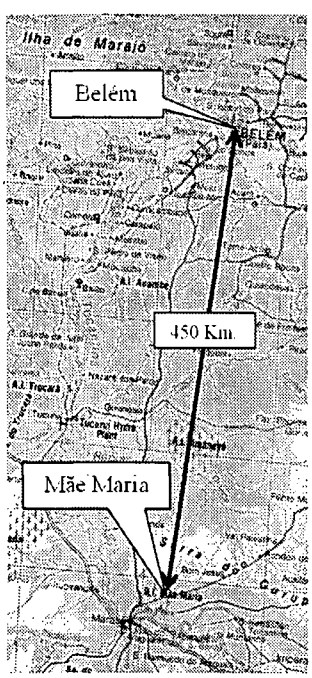
Elaboration d'un dictionnaire multimédia de la langue Parkatêjê, une langue Timbira de l'Amazonie brésilienne

Philippe Martin, Leopoldina Araújo
Université Paris 7 Denis Diderot, Universidade Federal do Pará
92, Ave. de France, 75013 Paris, FRANCE
Rua Avertano Rocha, 401, Cidade Velha, Belém, PARÁ, BRÉSIL
philippe.martin@linguist.jussieu.fr , leomaria@amazon.com.br

Résumé
Le Parkatêjê est une langue du groupe Timbira (Amazonie brésilienne), parlée dans un seul village par environ 400 locuteurs. La transcription des mythes de ce groupe a donné lieu à l'élaboration d'un lexique progressivement enrichi d'entrées et de rubriques nouvelles, d'abord sur papier, ensuite sur traitement de texte Word, enfin dans un programme de base de données Excel. Cette base de données utilise des liens qui permettent l'écoute d'exemples sonores liés aux entrées lexicales, ainsi qu'à leur analyse acoustique par le programme WinPitchPro.

Introduction



Le Parkatêjê appartient au groupe de langues Timbira (jê), parlées en Amazonie Brésilienne, qui comprend le Canela (Maranhão), l'Apinaye (Maranhão), le Kraho [Tocantins] et le Kayapo [Pará]. Elle est pratiquée aujourd'hui dans un seul village (Mãe Maria, Pará). La phonologie et la syntaxe du Parkatêjê ont été décrites par L. Araújo (1977), qui a également proposé un système orthographique.

Les Parkatêjê vivent au sud-est de l'État du Pará en Amazonie brésilienne – sur la *Terre Mãe Maria* qui se trouve à environ 450 Km au sud de Belém. Une étendue de 52.000 ha leur a été octroyée par le gouvernement Brésilien en 1943, et elle atteint actuellement 66.000 ha. La terre de ce peuple est en fait coupée par trois voies de pénétration, une route d'abord, (l'ancienne route de l'État, la PA70 devenue depuis 1999 une route fédérale asphaltée), la ligne haute tension de la centrale Hydroélectrique de Tucuruí, ainsi que le chemin de fer Carajás-Itaqui.

Le village Mãe Maria compte actuellement 400 personnes, soit une augmentation de 30% en huit ans. Les familles se répartissent en trois groupes : les Rõhôkatêjê (Parkatêjê), les Akrãtitakatêjê (montagne), et les Kyikatêjê (Maranhão). De nouveaux couples mixtes se sont formés, par mariage avec

des femmes brésiliennes.
Du point de vue linguistique, la première génération parle encore couramment la langue traditionnelle qui en fait le premier instrument de communication, et quelques individus du groupe Kyikatêjê ne comprennent que quelques mots de portugais et n'arrivent pas à s'exprimer dans cette langue.
La seconde génération utilise assez bien le portugais mais communique avec les anciens dans la langue traditionnelle. Les troisième et quatrième générations parlent presque exclusivement le portugais et entrent en contact de façon plus systématique avec la langue indigène par l'école.
Ayant décrit la langue dans le cadre d'une maîtrise en linguistique (1977), L. Araújo a été appelée par le chef du village Parkatêjê à transcrire les mythes. Les jeunes du village qui commençaient à voir la possibilité d'écrire leur langue lui ont demandé d'élaborer un lexique (présenté dans une thèse de doctorat en 1989), ce qui impliquait la constitution d'un système orthographique.
Cette première version du lexique, en plus de donner la traduction en portugais, spécifiait déjà la structure des mots et identifiait les champs sémantiques. En 1989 le chef Parkatêjê a proposé l'installation d'une école dans le village, et l'élaboration du dictionnaire a continué à s'enrichir de nombreux exemples.

La langue Parkatêjê

Morphologie

La plupart des racines du Parkatêjê sont monosyllabiques (*kwýr* "manioc", *kô* "eau"). Les rares exemples dissyllabiques ont l'accent placé sur la dernière syllabe (*ha'hêr* "mur", *py'ka* "terre"). Les racines peuvent se combiner avec des préfixes et des suffixes. Les préfixes sont non accentués, et ne modifient pas l'accentuation de la racine. D'autre part, les suffixes sont normalement accentués (sauf si la voyelle est réduite), et ne modifient pas non plus le patron accentuel de la racine. On obtient alors des mots contenant deux syllabes accentuées successives (*küm'xê're* "bacuri, un fruit régional", *ka'xê're* "étoile"). La collision d'accent est alors maintenue et n'entraîne pas un déplacement ou une réduction de l'accent comme ce serait le cas en anglais ou en français par exemple.

Phonologie

Le système phonologique possède onze consonnes et seize voyelles. Parmi les onze consonnes, deux n'existent pas en portugais : l'occlusive palatale /č/ et la glottale /ʔ/, peu fréquente. Les occlusives sont non voisées et les nasales sont toutes bilabiales ou dentales. La vibrante dentale /ɾ/ est non tendue, la fricative pharyngale /h/ peut se réaliser comme [h] , [y] , [ʒ] ou [č]. De même /y/ a trois réalisations [y] , [ỹ] et [ʒ].
Le système vocalique, dix voyelles orales et six nasales, a toutes les voyelles du portugais ainsi qu'une série de postérieures non arrondies (haute fermée et moyennes fermées et ouvertes). Les patrons syllabiques sont complexes, avec la possibilité de combinaison

d'occlusives et même de nasales bilabiales avec la vibrante dentale dans le onset et d'occlusives dans la coda.

Orthographe

Un système orthographique a été proposé en 1977, selon les principes établis pour les langues du groupe jê, par L. Araújo, en collaboration avec des membres de la communauté Parkatêjê. La correspondance phonème – graphème est donnée dans le tableau ci-dessous :

Phonème	Orthographe	Exemple
p	p	*Par* "arbre" *tep* "Poisson"
t	t	*Tir* "vivant" *pat* "soleil"
č	x	*Xà* "fruit" *kaxêr* "lune"
k	k	*Kô* "eau" *crik* "se taire"
ʔ	h	*Ahũre* "poule" *ihpa* "son bras"
m	m	*Mre* "crocodile" *mra* "pleurer"
n	n	*nã* "donc" *pam* "ara"
ɾ	r	*Rop* "jaguar" *hakăr* "nuit" *kra* "sanglier"
w	w	*Wai* "xamã" *kwyr* "manioc"
y	j/i	*Pjê* "terre" *paipai* "vomir"
h	h	*Hôr* "dormir" *pohy* "maïs"
a	a	*ipa* "mon bras" *kaha* "couleuvre"
a (nasal)	ã	*Irã* "fleur" *prãm* "avoir envie"
ɛ	e	*Tep* "Poisson" *te* "jambe"
e	ê	*Mpoxê* "fil" *inxe* "mère de EGO"
ẽ	ẽ	*kre kre* "manger"
i	i	*Kre* "fourmi"
ĩ	ĩ	*Krĩ* "village"
o	o	*Popok* "canne fine"
õ	õ	*Ato* "frère"
ɔ	ô	*Kô* "eau" *tôn* "tatou"
u	u	*Pur* "abatis"
ũ	ũ	*Kum* "fumée"
ɨ	y	*Mxoty* "pépin"
ɨ (nasal)	ỹ	*hỹ* "oui"
ɤ	ỳ	*kwỳr* "manioc"
ʌ	à	*Jàt* "patate douce"

Table 1 : Transcription orthographique du Parkatêjê

Élaboration d'un dictionnaire

Une version préliminaire d'un dictionnaire du Parkatêjê contenait environ 1000 mots. Le vocabulaire a été étendu et porté sur traitement de texte en utilisant pas moins de 4 variantes distinctes de la police Latin (Latin basic, Latin 1, Latin étendu A et Latin étendu additionnel) pour tenir compte des symboles orthographiques retenus dans la transcription originale (par exemple la lettre y surmontée d'un accent aigu ý). La saisie des entrées dans le programme Word® a donc demandé l'élaboration de macros pour éviter la recherche des symboles nécessaires sans passer par la table de caractères disponible dans Windows®.

POLICE	CARACTERES
Latin basique	A a E e I i O o U u Y y H h J j K k M m N n P p R r T t W w X x
Latin 1	À à Ã ã Ê ê Ô ô Õ õ
Latin étendu - A	Ĩ ĩ ũ Ũ
Latin étendu additionnel	Ẽ ẽ Ỳ ỳ Ỹ ỹ

Table 2 : Les 4 polices de caractères utilisées sous Word

Le projet de dictionnaire multimédia sous Excel® avait pour but, entre autres, d'enrichir le dictionnaire existant par de nombreuses rubriques et de le porter sur le programme de gestion de base de données.

Le problème principal du portage des fichiers Word vers Excel a trait à la multiplicité des polices (4 versions des polices Latin) dans chaque entrée lexicale du fichier sous Word. Il n'était pas rare d'avoir des entrées utilisant à la fois les 4 polices, définies pour chaque caractère orthographique de l'entrée considérée. Or le programme Excel ne permet pas cette possibilité : chaque case ne pouvant utiliser qu'une seule police à la fois. La solution qui a paru la plus simple a consisté à développer un programme de conversion des entrées sous Word en format Unicode, dont les implémentations sous différentes polices de caractères contiennent toutes les variations nécessaires des caractères Latin.

Un programme de parsing traitant les informations sous Word on donc été conçu et implémenté, en exportant directement dans Excel les données converties ventilées selon les rubriques existantes.

Ã 00C0	Ã 00C3	Ẽ 00CA	Ẽ 1EBC	Õ 00D5	Ô 00D4	Ũ 0168	Ĩ 0128	Ỳ 1EF2	Ỹ 1EF8
à 00E3	ã 00E3	ê 00EA	ẽ 1EBD	õ 00F5	ô 00F4	ũ 0169	ĩ 0129	ỳ 1EF3	ỹ 1EF9

Table 3 : Caractères Unicode utilisés sous Excel

Mise en format Excel®

Les champs de la version enrichie sous Excel comprennent :

La transcription orthographique
La transcription phonologique
Le nombre d'homonymes
La transcription phonétique
La classe grammaticale
L'analyse morphologique (préfixe + lexème + suffixe)
La traduction en portugais
La traduction en français
La traduction en anglais
La référence de la source écrite
Un exemple d'utilisation de l'entrée en contexte
Une analyse morphologique de l'exemple
Une traduction de l'exemple
Un commentaire du lexicographe
Un pointeur du fichier son de l'exemple

Un exemple d'affichage apparaît dans la figure ci-dessous.

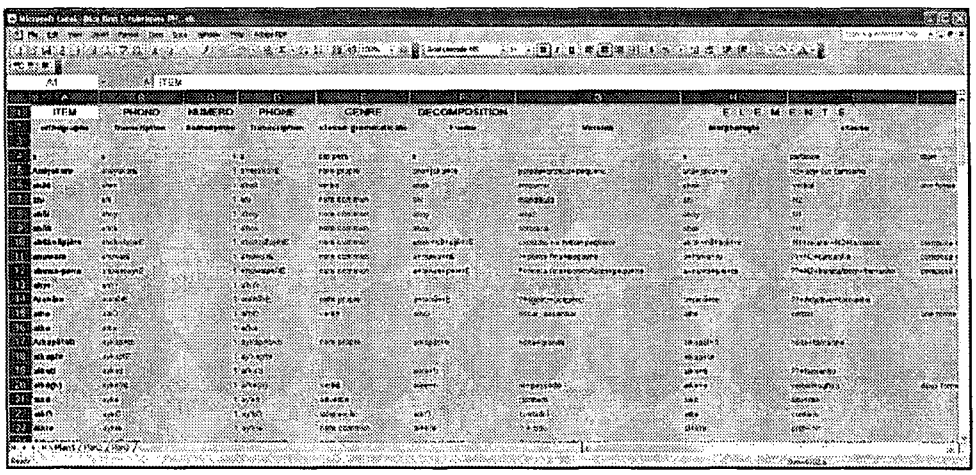


Figure 1 : Exemple d'affichage de la base de donnée du dictionnaire

Parmi les nombreuses possibilités qu'offre Excel, on a ajouté dans un premier temps, pour chaque entrée, un lien avec un fichier son (format wav) correspondant. Le but ultime était d'obtenir automatiquement une analyse acoustique (spectrogramme, courbes d'intensité et de fréquence fondamentale) de ces fichiers pour une description phonétique détaillée, particulièrement en ce qui concerne l'accentuation et l'intonation de la phrase, une caractéristique importante du fait que les nouvelles générations n'utilisent plus couramment la langue traditionnelle. Il est donc utile de pouvoir écouter les mots, des phrases et des petits textes oralisés en contexte.

La traduction en d'autres langues en plus que le portugais se justifie par différentes raisons : du point de vue académique, elle ouvre la possibilité de comparaison avec les donnés d'autres chercheurs ; du point de vue de la communauté, les jeunes qui étudient déjà aux niveaux secondaires moyen et supérieur s'intéressent à l'espagnol, à l'anglais et au français – il est donc important pour eux de voir leur langue traduite dans ces idiomes prestigieux.

L'affichage automatique de l'analyse acoustique des exemples sonores des entrées prises en contexte a demandé l'implémentation d'une fonction spécialisée dans le programme d'analyse phonétique retenu pour ce projet : WinPitchPro. Cette fonction permet d'afficher automatiquement l'analyse acoustique de l'entrée, par l'intermédiaire d'une fonction standard de WinPitchPro, permettant à l'utilisateur d'afficher l'analyse et d'écouter un fichier son en cliquant sur l'entrée désirée dans une table de données. La section correspondant à l'entrée sélectionnée s'affiche automatiquement, avec le segment sonore de l'entrée mise graphiquement en relief. Ce mode permet l'utilisation d'un seul fichier son, obtenu par concaténation des exemples pertinents en contexte, plutôt que le chargement d'un fichier distinct par entrée lexicale, entrées dont le nombre de plusieurs milliers rend difficile la gestion informatique.

Analyse phonétique et linguistique avec WinPitchPro

L'analyse phonétique des entrées lexicales du dictionnaire se fait avec le programme WinPitchPro. Ce programme permet également la transcription assistée, directement en format de données Unicode, des exemples sonores recueillis sur le terrain. Comme le montre la figure ci-dessous, la transcription des caractères non standard en police Latin est effectuée en cliquant sur une table de symbole définie par l'utilisateur, de manière à éviter des procédures d'entrée de caractères qui seraient autrement fastidieuses.
Une fois la transcription effectuée, une base de donnée élaborée automatiquement apparaît dans une table. Par un simple clic sur une entrée, la section sonore correspondante est automatiquement analysée acoustiquement. Cette base de donnée peut également être sauvegardée sous un format Excel et XML pour une utilisation ultérieure par d'autres programmes d'analyse.

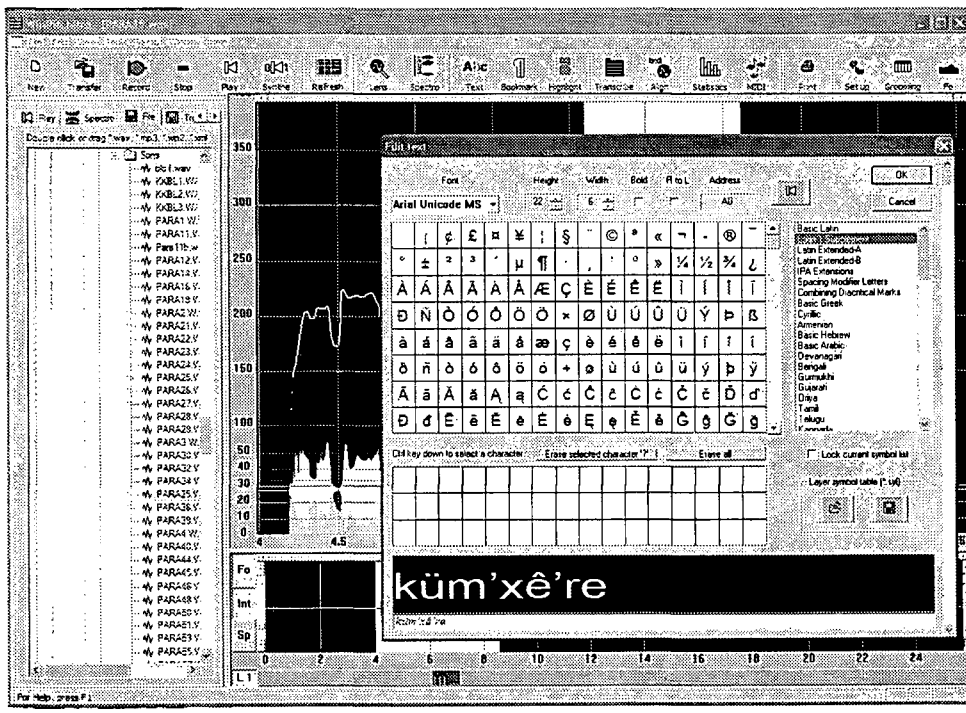


Figure 2 : Fenêtre de transcription, dans laquelle l'utilisateur peut élaborer une table de caractères spécialisés

Références

Araújo. L., 1989, *Aspectos da língua gavião-jê*, Rio de Janeiro, UFRJ, Tese de Doutorado.
Araújo, L., 1993, "Fonologia e grafia da língua da comunidade parkatêjê (timbira)" in: **Luci Seki** (org) *Lingüística indígena e educação na América Latina*, Campinas/SP, Editora da UNICAMP, pp 265-271.
"***Conhecendo nosso povo***", 1997, Comunidade Indígena Parkatêjê. Brasília: Ministério de Educação e Desportos, Belém:Secretaria de Estado de Educação.
Ferraz, I., 1983, "Os *parkatêjê das matas do Tocantins: a epopéia de um líder timbira*", São Paulo, USP, Dissertação de Mestrado.
Martin, Ph., 1998, Prosodie des langues romanes : analyse phonétique et phonologique. *Le français parlé*, GARS, Aix en Provence.
WinPitch, 1996, 2003, http://www.winpitch.com