



HAL
open science

Description du dictionnaire électronique des verbes simples du malgache

Joro Ny Aina Ranaivoarison

► **To cite this version:**

Joro Ny Aina Ranaivoarison. Description du dictionnaire électronique des verbes simples du malgache. Colloque Jeunes Chercheurs Sciences du Langage "Trace(s)", Oct 2015, Montpellier, France. . hal-01237502

HAL Id: hal-01237502

<https://hal.science/hal-01237502v1>

Submitted on 3 Dec 2015

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.

Introduction

L'exploration des textes est un moyen d'identification et de traçage des passages des hommes sur terre. La construction des dictionnaires électroniques est une des voies exploratoires que les chercheurs ont mises sur pied pour permettre une fouille rapide et puissante des données textuelles sous format électronique. Les travaux sur le français [2], notamment sur la syntaxe, ont inspiré de nombreux chercheurs dans le domaine des dépouillements informatisés de textes et les ont incité à construire des dictionnaires électroniques d'autres langues comme le DECO – dictionnaire électronique du coréen [5] et [7] – par exemple. Ce poster quant à lui décrit le dictionnaire électronique des verbes simples du malgache (DEMA-VS) qui est une langue austronésienne comme l'indonésien, le philippin et agglutinante comme le turc, le coréen. La méthode utilisée pour le construire y est présentée. La section consacrée aux résultats décrit ses caractéristiques générales, ses entrées, ses articles, et le dictionnaire des racines verbales.

Objectifs

L'objectif global est de construire un dictionnaire électronique du malgache (verbes, noms, adjectifs, adverbes, mots grammaticaux). La figure ci-dessous montre la structure générale de ce dictionnaire.

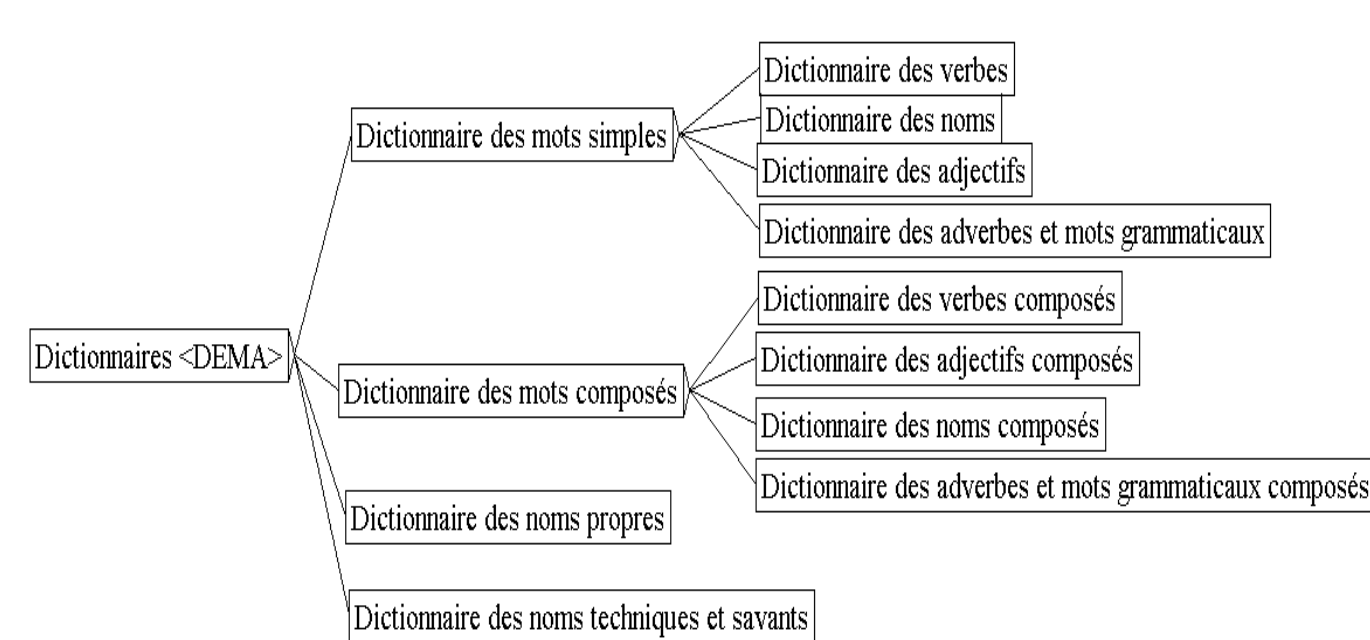


Fig 1. Structure générale du dictionnaire électronique (DEMA)

Les objectifs spécifiques sont de construire les dictionnaires des noms simples, des mots composés, des noms propres, techniques et savants.

Le dictionnaire des verbes simples [8] quant à lui, est disponible et téléchargeable sur Unitex [6] <http://igm.univ-mlv.fr/~unitex> sous licence LGPL.

Méthodes

La méthode utilise des notions linguistiques représentées par des codes comme V pour « verbe », f pour « futur », o pour « objectif », se base sur le format DELA [4] – un format utilisé pour la construction des dictionnaires électroniques du français – et s'inspire de [1], un analyseur morphologique à large couverture lexicale construit pour le coréen qui est une langue agglutinante comme le malgache.

En pratique, il s'agit de construire des tables de verbes avec des propriétés qui sont les variations morphologiques des verbes, les modes de combinaisons des affixes afin de les coder et de construire des transducteurs qui opèrent directement sur le dictionnaire pour construire automatiquement un dictionnaire qui permet l'identification automatique des mots.

DEMA-VS

Caractéristiques générales

Le DEMA-VS est un dictionnaire morphologique et flexionnel à large couverture lexicale qui peut être facilement mis à jour par les linguistes. Ses entrées sont les radicaux verbaux. Sa taille actuellement est de 207 Ko et ses articles sont composées de codes informatiques (cf. [10]). Ces codes informatiques sont les codes de classes affixales (455), de classes radicales (154) et de modèles de conjugaison (196) qui est un niveau de granularité intermédiaire entre les classes et les groupes de conjugaison (3) [9]. La structure générale du dictionnaire est « entrée,Vclasseradicale+classeaffixale+groupe+modèle+numclasse ». La figure ci-dessous montre l'architecture générale de ce dictionnaire.

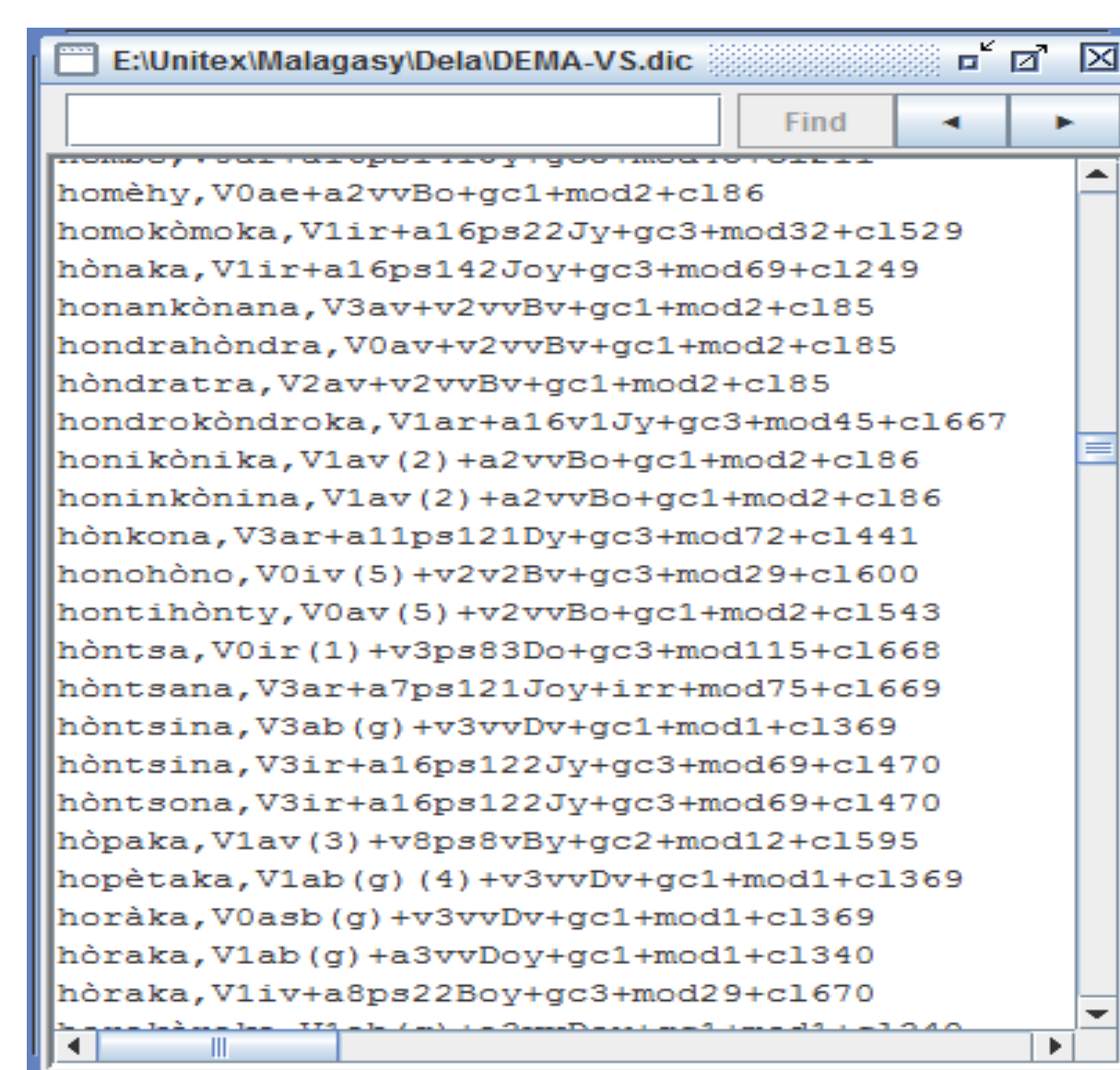


Fig. 2. Architecture du DEMA-VS

Entrées

Le dictionnaire comporte un peu plus de 3000 entrées. Une fois qu'une entrée est introduite dans le dictionnaire, cela permet de reconnaître automatiquement dans un texte toutes les formes verbales conjuguées de cette entrée.

Pour *hòraka* « A. Applaudir, acclamer, huer, pousser des cris de réprobation. B. Coasser. » par exemple, une fois l'entrée correspondante introduite dans le dictionnaire électronique, toutes les formes verbales de ce radical sont reconnues automatiquement par l'analyseur morphologique d'Unitex telles que *nìhoraka*, *nahòraka*, *nohorahina*, *voahòraka*, *tafahòraka* au passé, *mìhoraka*, *mahòraka*, *horahina*, *voahòraka*, *tafahòraka* au présent, et *hìhoraka*, *hahòraka*, *hovoahòraka*, *hotafahòraka* au futur.

Chaque ligne du dictionnaire contient une entrée constituée d'un radical, d'un séparateur (la virgule) et de codes représentant les propriétés morphologiques du radical.

Codes

Les codes sont séparés par le signe « + » (convention d'Unitex).

- catégorie grammaticale des verbes (V) suivi des codes de la variation morphologique (classe radicale)
- combinaison des affixes (classe affixale)
- groupe de combinaison (gc)
- modèle de conjugaison (mod)
- classe de conjugaison (cl).

À l'aide de ces codes et des transducteurs préalablement construits, le compilateur d'Unitex permet de générer le dictionnaire électronique des racines des verbes (DEMA-VSflx).

Dictionnaire des racines verbales (DEMA-VSflx)

Le DEMA-VSflx est le dictionnaire électronique des racines verbales qui apparaissent dans les formes fléchies des verbes. Chaque ligne du DEMA-VSflx est composée de : « la forme de surface d'un radical verbal, la forme canonique de celui-ci. la catégorie grammaticale des verbes + les codes de combinaison des affixes + les codes de groupe de conjugaison + les codes de modèle de conjugaison + les codes de classe de conjugaison + les affixes spécifiques pour une forme de surface donnée ». Ci-dessous, un aperçu de ce dictionnaire est fourni.

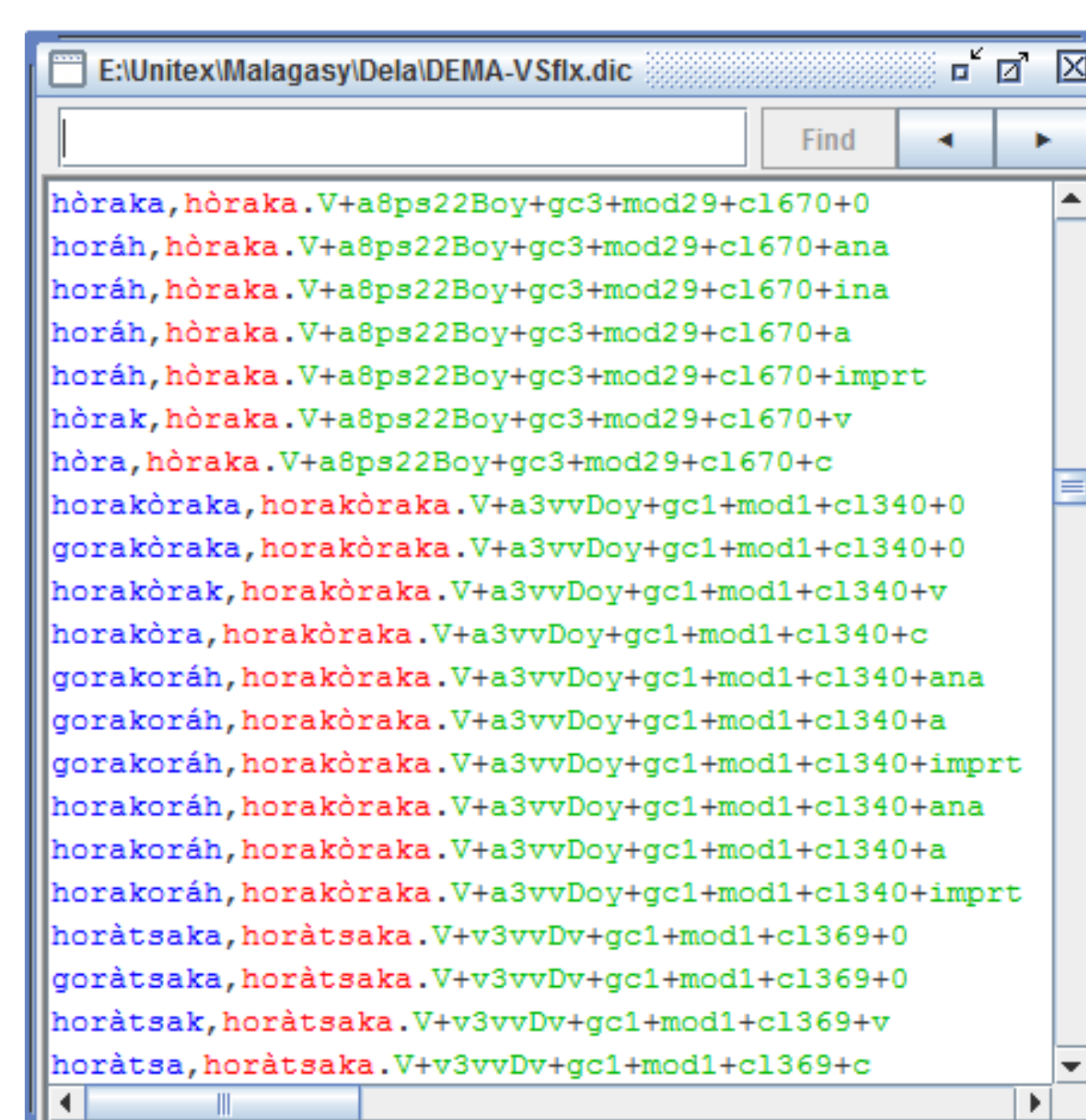


Fig 3. Architecture du DEMA-VSflx

Conclusions

La construction de tels dictionnaires pour une langue permet de garder ses traces et enrichit les possibilités de fouille documentaire la concernant. La documentation automatique (voir. [3]) est courante de nos jours, nous pensons que la construction des autres ressources linguistiques du malgache (noms, adjectifs, adverbes, mots grammaticaux) enrichira les applications existantes et ouvrira celle-ci à des applications utilisables par le grand public.

Bibliographie

- [1] Berlocher, I., Huh, H.G., Laporte, É., Nam, J.S. (2006). Morphological annotation of Korean with Directly Maintainable Resources, Poster. In: *Proceeding of LREC*. Genoa.
- [2] GROSS, M. (1975). *Méthode en syntaxe. Le régime des constructions complétives*. Paris: Hermann.
- [3] Laporte, E. (2000). Mots et niveau lexical. Jean-Marie Pierrel. Ingénierie des langues, Hermès, pp.25-49, Informatique et systèmes d'information.
- [4] Laporte, E. (1993). Separating entries in electronic dictionaries of French. In: J. Darski and Z. Vetulani, (eds.), *Sprache - Kommunikation - Informatik, Akten des 26. Linguistischen Kolloquiums, Poznan, Tübingen: Max Niemeyer, p. 173-179*.
- [5] Nam, J. S. (1994). Constitution d'un lexique électronique des noms simples en coréen. In: J. Labelle, *Lexique-grammaires comparés et traitements automatiques*. Montréal, Service des Publications, UQAM.
- [6] Paumier, S. (2015). *Unitex 3.1 Bêta Manuel d'utilisation*. Université Paris-Est Marne-la-Vallée.
- [7] Paumier, S., Nam, J. S. (2014). Un système de dictionnaire de mots simples du coréen. In: Fryni Kakoyianni-Doa, *Penser le Lexique-Grammaire. Perspectives actuelles*, Honoré Champion, pp.481-490, Collection Colloques, congrès et conférences. Sciences du Langage, histoire de la langue et des dictionnaires. 30th International Conference on Lexis and Grammar (Nicosia, Cyprus, 2011), 978-2-7453-2512-9.
- [8] Ranaivoarison, J. N. A., Laporte, E., Ralalaoherivony, B. S. (2013). Formalization of Malagasy conjugation. In: *Language and Technology Conference*. Poznań, Poland. pp.457-462.
- [9] Ranaivoarison, J. N. A. (2014). *Modélisation de la morphosyntaxe du malgache. Construction d'un dictionnaire électronique des verbes simples*. Thèse de doctorat. Université d'Antananarivo.
- [10] Silberztein, M. (1993). *Dictionnaires électroniques et analyse automatique de textes. Le système Intex*. Paris: Masson.