

Description formelle des termes malgaches

RAVELONJATOVO Tantely Harinjaka
Centre Interdisciplinaire de Recherche Appliquée
au Malgache (CIRAM)
Université d'Antananarivo - Madagascar

PLAN

- INTRODUCTION
- CADRE METHODOLOGIQUE
- LE CORPUS DE TRAVAIL
- DESCRIPTION FORMELLE
- CONCLUSION

INTRODUCTION

- OBJECTIFS
 - généraux (de la thèse)
 - spécifiques (de la description formelle)
- PROBLEMATIQUE
 - générale (de la thèse)
 - spécifique (actuelle : description formelle)
- HYPOTHESES
 - hypothèses générales
 - hypothèses spécifiques

INTRODUCTION

OBJECTIFS

Généraux : Étudier la création des termes malgaches. Cas du domaine de l'environnement.

Spécifiques : Validation de la formation dans le corpus. Cas du phrasème de type NV

INTRODUCTION

- PROBLÉMATIQUES

- générale : comment se forment les termes en langue malgache?

Français :

Ex : extraction automatique possible

composition savante: –logie, etc...

Malgache :

Ex : phrasème NV

- spécifique : comment valider la formation phrasème de type NV?

INTRODUCTION

- HYPOTHESES
 - Générales : hypothèses de la thèse
 - Spécifiques : hypothèses par rapport à la validation sur corpus

INTRODUCTION

HYPOTHESES

- Hypothèses générales

Comment se créent les termes en malgache?

- **Morphologiquement**, un terme malgache se forme généralement en respectant les règles de formation lexicale de la langue malgache générale
- Sémantiquement, le sens d'un terme se rapporte à son domaine

INTRODUCTION

HYPOTHESES

- Hypothèses spécifiques:
 - comment valider la formation du phrasème de type NV?
 - en le décrivant de manière formelle et (systématique)
 - en le mettant dans le contexte (le corpus).

Cadre méthodologique

- Linguistique de corpus (Sinclair, 2004)
 - Corpus :
 - Attestation dans le corpus (exemples avérés)
 - Attestation vs illustration
 - Attestation vs introspection
 - *Data-driven approach* : approche guidée/dirigée par les données (Williams, 1999)
 - Élaboration d'hypothèses à partir du corpus
- Terminologie textuelle [Slodzian et Bourrigault, 1999]
 - Terminologie descriptive
 - Terminologie basée sur les textes

Cadre méthodologique

- Constitution d'un corpus spécialisé (TONTONA*)
 - Constitué et organisé selon :
 - des critères linguistiques
 - des critères extralinguistiques
- Constitution d'un échantillon de termes
- Etude de leur forme et de leur formation
- Etude de termes par rapport aux situations de communication
- Validation des hypothèses sur corpus

Le corpus de travail

- Corpus : brut, monolingue
- Langue : malgache écrite
- Textes : originaux et traduits
- Auteurs : locuteurs natifs
- Étude : synchronique
- Date de publication des textes : 1990 -2005
(Programme National pour l'Environnement)
- Domaine : environnement

Le corpus de travail

Niveau	Typologie	type		
A	Typologie de langue	Langue générale vs Langue spécialisée		
B	Typologie de discours	scientifique	juridique	Journalistique
C	Typologie de documents	Articles scientifiques et mémoires	Textes de droit, chartes, contrats de transfert de gestion.	Articles de journaux et autres articles

Le corpus de travail

Répartition de textes dans les situations de communication

TONTONA			TOTAL
CORPUS			
1 - Scientifique	2 – Juridique	3 – journalistique	3 sous- corpus
56 textes 33 auteurs	09 textes 09 auteurs	15 textes 11 auteurs	80 textes 52 auteurs
178324 mots	384213 mots	8812 mots	571349 mots

DESCRIPTION FORMELLE

- Objectifs : validation des hypothèses sur la formation des termes en malgache par
 - description des termes pour la validation des hypothèse dans le corpus
 - Observation contextuelle dans le corpus.
- Problématique actuelle: comment décrire ces termes avec Unitex avec un minimum de ressources linguistiques?
 - Création de dictionnaire ou construction des graphes?
- Cas étudié : phrasème de type NV

DESCRIPTION FORMELLE

- Cas du phrasème de type NV
 - Définition :
 - “Un **phrasème** de la langue L est une expression multilexémique de L qui ne peut pas [...] être produite, à partir d’une situation donnée ou d’un sens donné, selon un dictionnaire de mots de L et à partir des règles générales standard de L” (Mel’čuk 1993:83)
 - “le phrasème est un complexe phraséologisé de signes [Mel’čuk, 1997]
 - Le phrasème terminologique est à la fois
 - Syntagmatique [COLLET, 2004]
 - Complexe [Cabre, 1998] [L’Homme, 2005]
 - Le phrasème est liée à la phraséologie (**Svensson, 2004**)
 - Autres types phrasèmes : NV, VN, NN, etc.

DESCRIPTION FORMELLE

- Cas du phrasème de type NV
- Il existe trois types de phrasème NV (selon la voix)
 - NVcirc (Verbe : voix circonstancielle). Si V est passif à circumfixe x...ana, N est un complément circonstanciel (indirect).
 - NVobj (Verbe: voix objective) Si V est passif à suffixe -ina, -ana, à circumfixe ha...ina ou préfixe voa-, N est un complément d'objet direct. Verbe composé possible
 - NVage (verbe:voix agentive). Si V est actif (à préfixe), N est sujet, agent.

DESCRIPTION FORMELLE

- Cas du phrasème de type NV
- NVcirc
 - Ex : tontolo iainana (monde où l'on vit > environnement)
 - Circonstanciel : **lainana** ny tontolo (le monde est vécu)
 - Agentif : **miaina** ao anaty tontolo i Luc (Luc vit dans un monde)
- NVage
 - Ex : biby **mampinono** (animal allaitant>mammifère)
 - Agentif : **mampinono** ny biby (l'animal allaite)
 - Objectif : **ampinonoana** ny biby (l'animal est allaité)
- NVobj
 - Ex: faritra arov**ana** (zone protégée)
 - Objectif : arov**ana** ny faritra (la zone est protégée)
 - Agentif : **miaro** ny faritra i Luc (Luc protège la zone)

DESCRIPTION FORMELLE

- Principes de la description
 - Analyse en éléments des termes
 - Description de la variation des termes et des éléments constitutants
- Hypothèses
 - Termes : $[(\text{base})^* + (\text{affixe})]^*$
 - Base : radical (domaine général)
- Niveaux abordés
 - morphologie flexionnelle
 - morphologie dérivationnelle
 - phonétique combinatoire

DESCRIPTION FORMELLE

HYPOTHESES

- Hypothèses générales:

Morphologiquement,

- La création ou formation des termes en langue malgache repose sur le terme simple :
 - base+affixe
- La formation de terme complexe est secondaire.
 - (Base+(affixe)*)*
 - faute de terme simple,
 - En attendant le terme définitif

Remarque : les termes provisoires peuvent devenir termes définitifs

DESCRIPTION FORMELLE

- Hypothèses
 - Termes = [(base)*+(affixe)]*
 - Ex : *tontolo* (base) *iaianana* (base+affixe) (Environnement : base+affixe)
 - Base = radical (domaine général)
 - Ex : *tontolo* : *izao tontolo izao* (le monde),
fanatontolona/fanontoloana (mondialisation)
 - *iaianana* : *miaina* (vit), *fiaina* (la vie).
 - Exceptions
 - Emprunt : *zenetika* (génétique)
 - Formation savante : *Vohemarensis* < *Vohémar* (nom de lieu) + *ensis* (base latine)
 - Sigle et Acronyme : *Tanteza* > *fiTantanana hateza* ny ala
(*G*estion *l*ocale *s*écurisée)

DESCRIPTION FORMELLE

LES NIVEAUX ABORDÉS

- morphologie flexionnelle

Affixation > conjugaison

- morphologie dérivationnelle

Affixation > formation lexicale

- phonétique combinatoire

variation combinatoire

DESCRIPTION FORMELLE

- Morphologie flexionnelle

Français (**suffixe**): conjugaison (mode, temps, personne, etc...), genre.

Malgache : (**préfixe, suffixe, circumfixe**)

- mode : indicatif, impératif (**suffixe**)

Ex : matory-Ø (dort), matoria**a** (dors)

- temps : présent, passé, futur (**préfixe**)

Ex : **m**iaina (vit), **n**iaina (vivait), **h**iaina (vivra)

- voix : agentif, statif, agissif, objectif, instrumentative, applicative, circonstanciel, causatif, réciproque (**affixes**).

Remarque : Il n'y a pas de genre en malgache.

DESCRIPTION FORMELLE

Morphologie flexionnelle (phrasème NV)

VERBE

mode : indicatif

voix :

agentive : *loza mitatao* (danger sur la tête> menace), *zava-manana aina* (chose avoir vie > être vivant)

circonstancielle : *tontolo iainana* (monde où l'on vit>environnement), *faritra ivelomana* (zone où l'on survit> zone de survie)

objective : *faritra arovana* (zone protégée), *ala voajanahary* (forêt naturelle), *zava-boahary* (chose créée>nature)

CONSÉQUENCES :

Variation de l'affixe + variation du radical

Ex de variation de l'affixe : *i...ana>i...na* (*tontolo iainana*) ou rad = aina, voa>boa (*zava-boahary*) ou rad = zavatra et hary,

Ex de variation du radical : *velona>velom* (*faritra ivelomana*) ou rad = velona; aro>arova (*faritra arovana*) ou rad = aro

DESCRIPTION FORMELLE

Description de la morphologie flexionnelle
(phrasème NV)

- Ressources nécessaires
 - Liste des radicaux (+ variantes)
 - Mode variation flexionnelle des radicaux
 - Ex : velona > velom, aro>arov
 - Liste des affixes flexionnelles (+ variantes)
 - Mode de variation des affixes flexionnelles
 - Ex : i...ana>i...na.

DESCRIPTION FORMELLE

- Morphologie dérivationnelle
 - Français :
 - Dérivation : propre, affixation, impropre
 - Affixe : préfixe, affixe
 - Malgache :
 - Dérivation
 - Affixes : suffixe, affixe, circumfixe

DESCRIPTION FORMELLE

Morphologie dérivationnelle (phrasème NV)

- Affixation : 0
- Suffixation

Ex : *harena* < *hary* + *ana* (richesse) dans
 <*harena voajanahary* (richesse naturelle),
 toerana < *toetra* + *ana* (endroit, place) dans
 toerana arovana (endroit protégé)

- Circumfixation : 0

DESCRIPTION FORMELLE

Description de la morphologie
derivationnelle (phrasème NV)

CONSÉQUENCES

Dérivation affixale : affixe+base>variation de
l'affixe et du radical

Variation de l'affixe : **ana**>**na** dans hare**na**

Variation du radical : **hary**>**hare** dans **hare**na

toetr>**toer** dans **toer**ana

DESCRIPTION FORMELLE

- Description de la morphologie dérivationnelle
 - Ressources nécessaires
 - Liste des radicaux (+ variantes dérivationnelles)
 - Mode de variation dérivationnelle des radicaux
 - Liste des affixes dérivationnelles (+ variantes)
 - Mode de variation dérivationnelle des affixes

DESCRIPTION FORMELLE

- Phonétique combinatoire
 - Français : élision, etc...(deux voyelles successives : premier mot et deuxième mot, voyelle et h muet)
 - L'hiver < le hiver
 - d'amour > de amour
 - Malgache :
 - Variation < ajout de substituts personnels dépendants
 - ko, (ma, mon, mes), nao (ta, ton, tes,), ny(sa, son, ses,), nay « exclusif » (notre, nos), ntsika « inclusif » (notre, nos), ny (leur, leur)
 - Ex : satroko < satroka + ko (mon chapeau), tongony < tongotra (ses pieds)
 - Variation combinatoire < composition
 - Modification au niveaux des frontières des premier et deuxième mot.

DESCRIPTION FORMELLE

Description de la phonétique combinatoire
(phrasème NV)

- Ajout de substituts : 0
- Composition
 - *Loharano* < *loha* + *rano* (tête eau>source)
Aucune modification morphologique
 - *Voajanahary* < *voatra*+*zanahary* (fabriquer + Dieu>naturel)
 - *Loharanon-karena* < *loha* + *rano* + *na* + *harena*
(source de richesses > ressources)
 - *Valanjavaboary* < *vala*+*na*+*zavatra*+*voary* (parc naturel)

DESCRIPTION FORMELLE

Phonétique combinatoire (phrasème NV)

Composition : base + base > modification des frontières

- *voa***tra** > *voa-* ; *zanahary* > *j*anahary dans *voa***j**anahary (naturel)

- *na*>*n-* ou *n* ; *zavatra* > *j*ava, *voary*>*b*oary dans *valan***j**ava**b**oary (parc naturel)

DESCRIPTION FORMELLE

- Description de la phonétique (variation) combinatoire (phrasème NV)
- Ressources nécessaires
 - Liste des radicaux (+ variantes combinatoires)
 - *Voatra*>*voa*, *zanahary*>*janahary* dans *voajanahary*
 - *Zavatra*>*zava*, *voary*>*boary* dans *zava-boary*
 - Mode de variation combinatoire
 - Selon les traits phonétiques (ex : z et j sont des consonnes dentales, v et b sont des labiales)

DESCRIPTION FORMELLE

Récapitulation des ressources

- option 1 : liste des mots et leurs variantes
 - affixes
 - radicaux
- option 2 : liste des mots et description des modes de variations
 - variation flexionnelle
 - variation combinatoire

DESCRIPTION FORMELLE

Récapitulation des ressources

- Liste des radicaux (+ variantes flexionnelles)
 - toetra > toer, aro>arov,
- Liste des radicaux (+ variantes combinatoires)
 - voatra > voa,boa
- Liste des affixes (+variantes flexionnelles)
 - i...ana>i...na,
- Liste des affixes (+variantes combinatoires)
 - voa>boa, i...ana>i...an|i...am,

DESCRIPTION FORMELLE

Récapitulation des ressources

- Deux variations (langue agglutinante)
 - Variation combinatoire (RAJAONA, 2004)
 - Ex1 : *voajanahary* < *voatra*+*zanahary* (naturel)
Où *voatra* (*construire*) est un verbe et *zanahary* (Dieu) un nom
 - Variation morphophonologique (RAJAONA, 2004)
 - Ex : *maniry* (pousse) < man - *tsiry* (agentif) dans *zava-maniry* (plante)
 - $n+ts > n$ vs $n+ts > nts | n-ts$ (variation combinatoire). Ex :
olona+tsotra > *olon-tsotra* (tsotra : simple, ordinaire, olona : personne)
- Remarque : quelquefois, variation morphophonologique = variation combinatoire > ambiguïté

DESCRIPTION AVEC UNITEX

Problématique actuelle: comment décrire ces phrasèmes avec Unitex?

- Construction des graphes

- Solutions 1 : utilisation des expressions exactes ou expressions régulières.
- Solutions 2 : utilisation de dictionnaires
- Solutions 3 : ajout d'informations linguistiques

DESCRIPTION AVEC UNITEX

Solution 1 : expressions rationnelles

Ressources utilisées : préfixes.

- phrasème NVcirc

<MOT>.<MOT><<^i.*(ana|na|an|am\$)

pour rechercher les termes *tontolo iainana* et *faritra ivelomana*

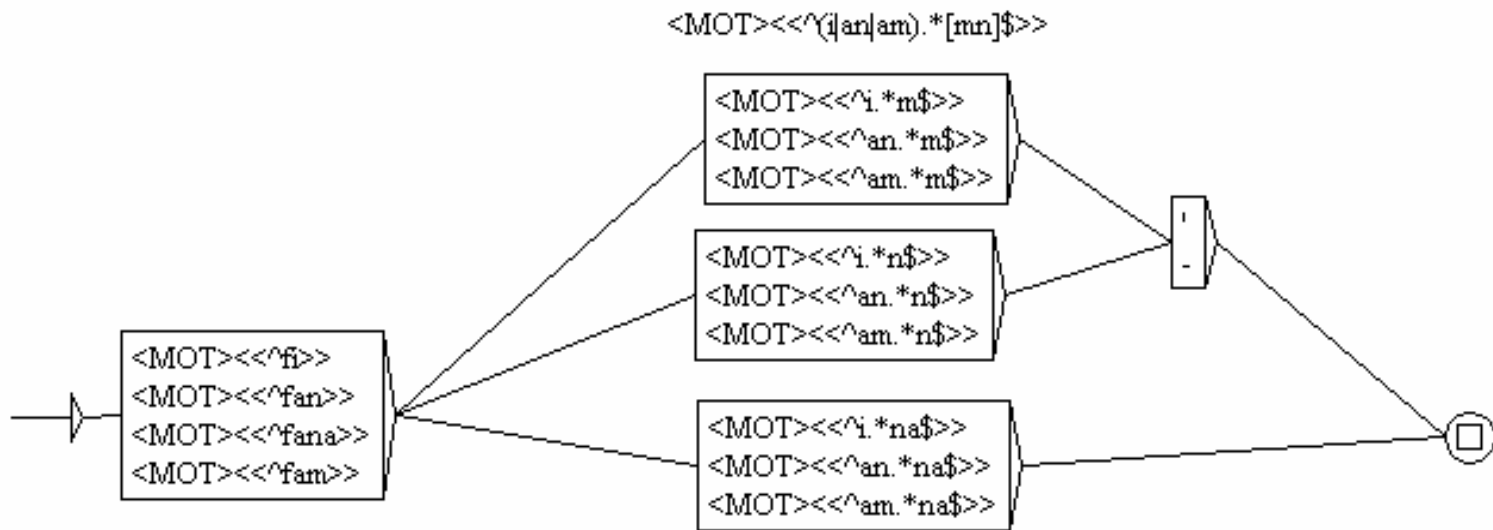
- phrasème NVage

<MOT>.<MOT><<^man|mi|ma>> pour rechercher les termes *biby mandady*, *zava-maniry*, *loza mitatao*, etc...

- phrasème NVobj **<MOT>.<MOT><<^(voa|ha).*?(ina|ana)>>**
pour rechercher les termes *faritra arovana*, *toerana voaro*, *harena azo havaozina*

DESCRIPTION AVEC UNITEX

Graphe construit avec des expressions régulières (pour NVcirc)



DESCRIPTION AVEC UNITEX

- Résultats
 - erreurs
 - Beaucoup de bruits
- Commentaires
 - Abondance de bruits (séquences inutiles)
 - Beaucoup de silence (termes non trouvés)
 - Solution précaire (bricolage)
- Donc : abandon

DESCRIPTION AVEC UNITEK

- Solutions 2 : utilisation des dictionnaires
 - Ressources utilisées : dictionnaire Brill, affixes
 - Formattage du dictionnaire Brill_malgache (CIRAM, 2000)
 - Présentation du dictionnaire
 - aboabo,.SBC (ou SBC = substantif)
 - adala,.ADJ (ou ADJ = adjectif)
 - adala,.SBC
 - adika,.VER (ou VER = verbe)
 - Requête
 - <SBC>.<VER><<^i(ana|an|am)?>> pour NVcirc (Ex : *tontolo iainana*)
 - <SBC>.<VER><<^(man|mi|man)>> pour NVage (Ex : *zava-maniry, biby mampinono, biby mandady*)
 - <SBC>.<VER><<^(voa|ha)(ana|an|am|ina|na)>> pour NVobj (Ex : *faritra arovana, ala voajanahary, harena azo havaozina.*)

DESCRIPTION AVEC UNITEX

- Résultats (utilisation du dictionnaire Brill)
 - erreurs
 - beaucoup de bruits
 - beaucoup de termes non trouvés
- commentaire :
 - Insuffisance d'informations linguistiques dans le dictionnaire
 - Catégorie grammaticale uniquement
 - Il existe très peu de ressources en langue malgache
 - Dictionnaire de 8000 mots fléchis avec étiquettes morphosyntaxiques
 - Liste des 3900 radicaux bruts (sans étiquettes)

DESCRIPTION AVEC UNITEX

Solution 3 : ajout d'informations linguistiques

- Option 1

 - = dictionnaire électronique du malgache

 - = une thèse

- Option 2

 - graphe d'étiquetage

DESCRIPTION AVEC UNITEX

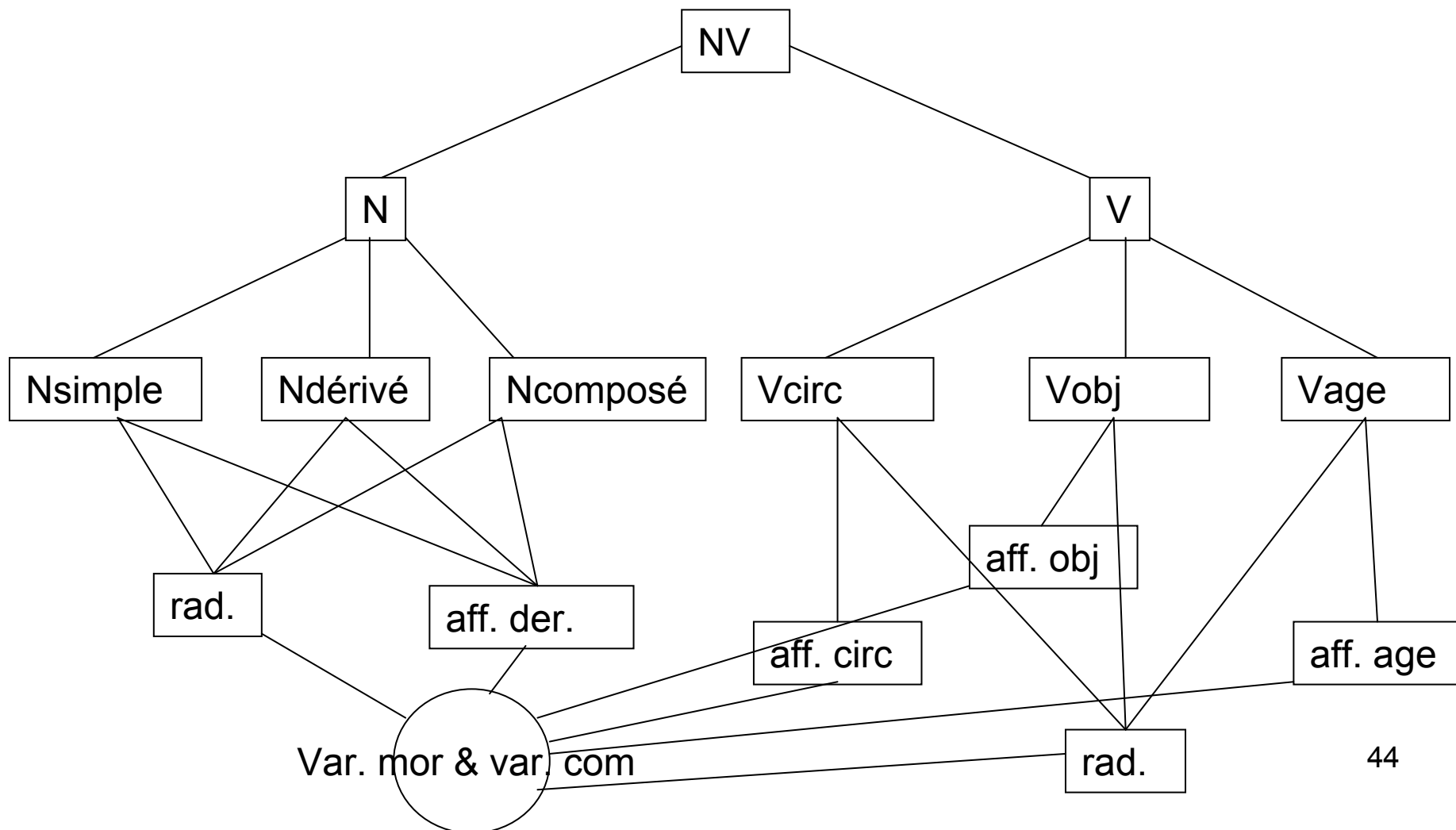
Option 2 : graphe d'étiquetage

- graphes de prétraitement
- options supplémentaires :
 - *Analyse this language char by char*
 - *Enable morphological use of space (new)*
- recherche d'expressions avec l'étiquette <...>
- génération possible de dictionnaire (graphe dictionnaire)

A voir Paumier, 2006, Unitex 1.2. Manuel d'utilisation.

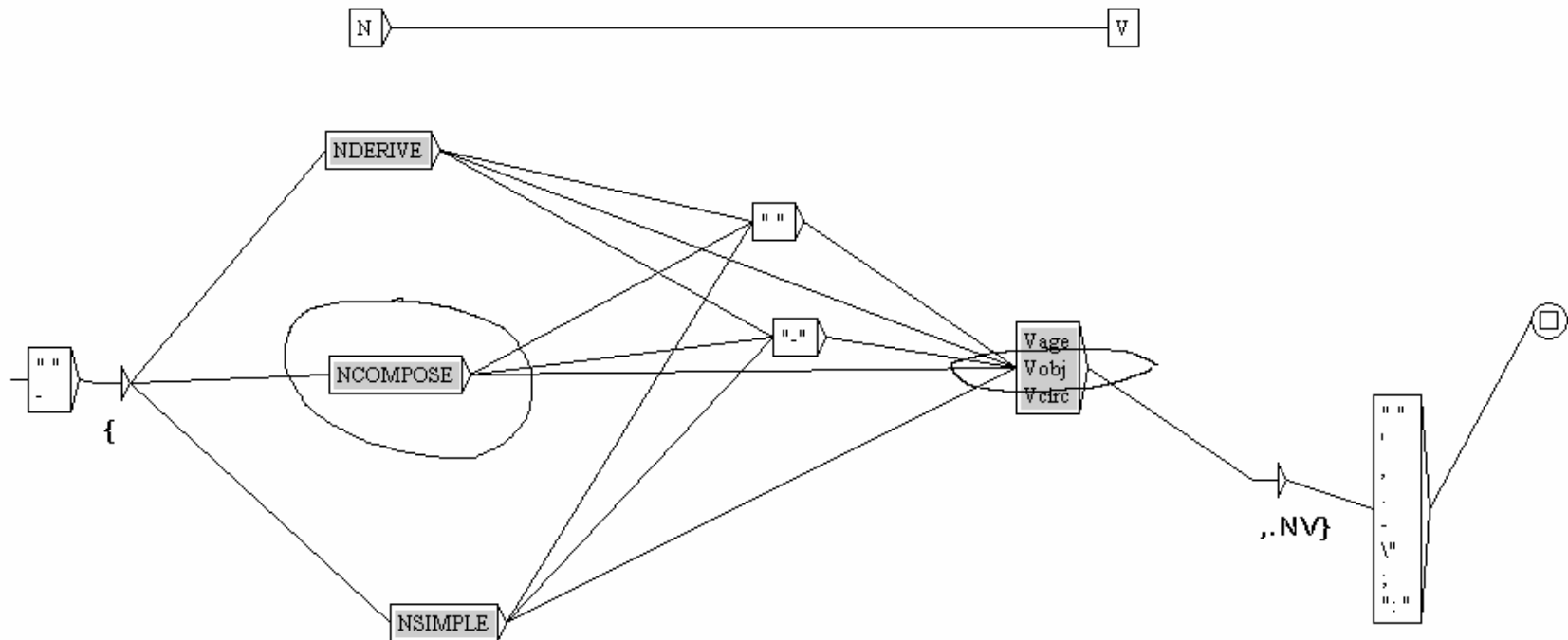
DESCRIPTION AVEC UNITEX

Les graphes de description des phrasèmes NV



DESCRIPTION AVEC UNITEX

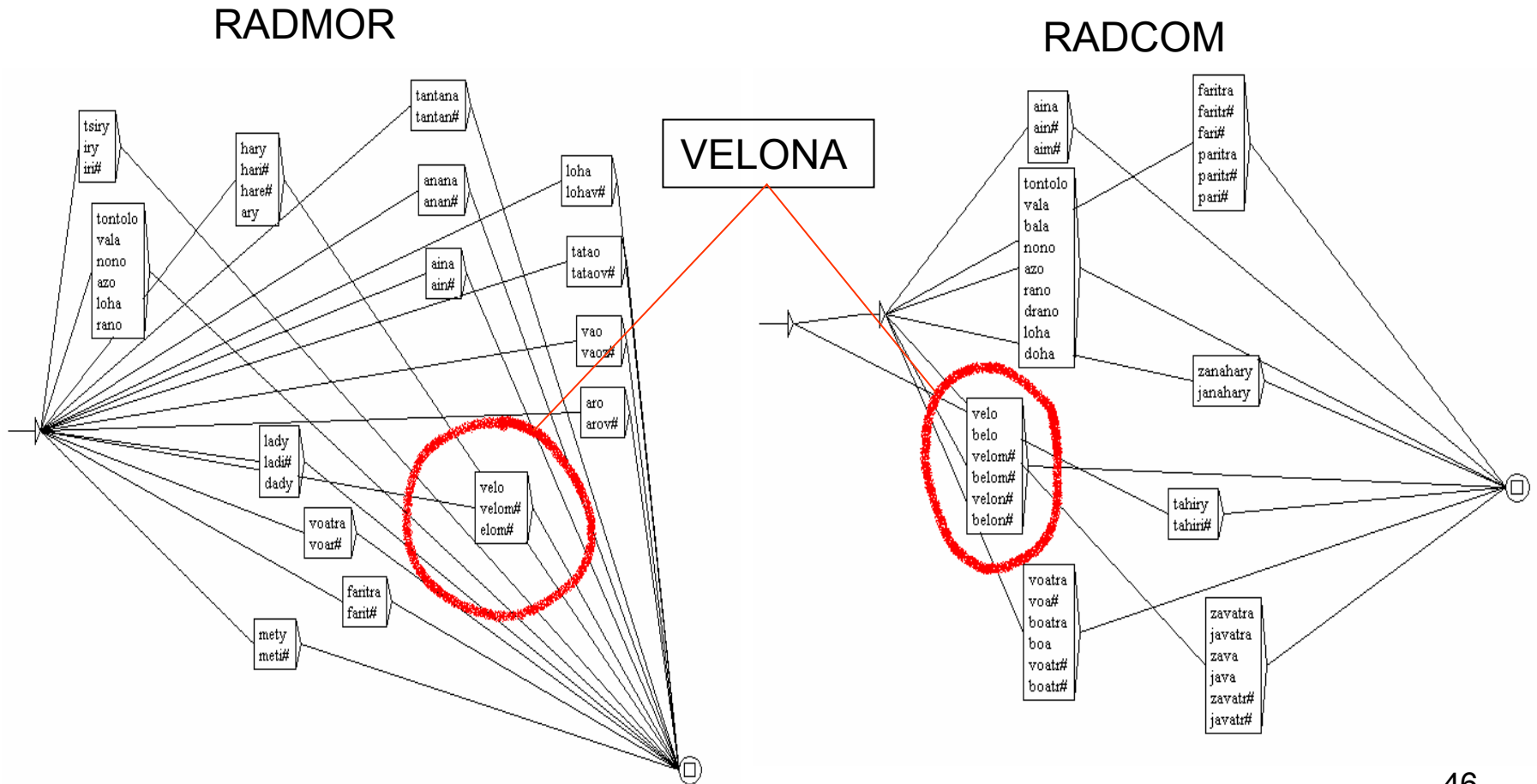
Graphe principal : phrasemeNV



DESCRIPTION AVEC UNITEX

(Variations morphologiques et combinatoires des radicaux)

RADMOR.grf et RADCOM.grf



DESCRIPTION AVEC UNITEX

- Test et résultats
 - Moitié des termes sont trouvés
 - Plusieurs expressions non souhaitées trouvés (bruits)
 - Moitié des termes non trouvés (silence)
- Solutions
 - Correction des bugs (M. Paumier)
 - Révision linguistique
 - Réfection des graphes (stt nom composés)

CONCLUSION

- Description des termes (phrasèmeNV)
 - Phrasèmes NV : base (N) + base (V)
 - Base (N) : mots simples ou mots construits
 - Mots construits : simples ou composés
 - Base (V) : verbes simples ou groupes verbales
 - Verbes simples : base + affixe
 - Groupe verbal : V simples + V simple ou base (N)

CONCLUSION

- Les variations
 - Les variations morphologiques
 - radicaux
 - affixes
 - Variations combinatoires
 - radicaux
 - affixes

PERSPECTIVES

Avec Unitex :

- Élaborer un dictionnaire malgache
- Décrire les modes de variations morphologique et combinatoire

Merci pour votre attention