

hyph-malagasy

Guide d'utilisation

Installation, compilation et contribution

Auteur : RALAHADY Bruno Bakys

ralahadybru@yahoo.fr

2 septembre 2026

Avant de commencer

Ce guide explique comment installer, compiler et faire évoluer le paquet hyph-malagasy au quotidien. Pour les décisions linguistiques et les résultats détaillés de validation (exécution réelle pdfTeX/LuaTeX), voir le document séparé *hyph-malagasy-documentation.pdf* (methodologie).

Version couverte par ce guide : 0.3.0-dev. Le corpus (29 mots) a été validé par un locuteur natif le 2 septembre 2026 ; la validation par l'Academia Malgache et l'intégration officielle à hyph-utf8 restent à faire.

1. Installation

Copier les fichiers du dossier tex/ dans un arbre texmf local, en respectant la structure TDS (TeX Directory Structure) :

```
~/texmf/tex/generic/hyph-malagasy/hyph-mg.tex
~/texmf/tex/generic/hyph-malagasy/loadhyph-mg.tex
```

Puis régénérer la base de données des noms de fichiers de votre distribution :

```
texhash # ou mktexlsr selon la distribution
```

Important : l'intégration complète au mécanisme `\uselanguage` de hyph-utf8 (chargement automatique via un nom de langue enregistré dans TeX Live) n'a pas encore été mise en place officiellement. En attendant, utilisez l'une des deux méthodes de test décrites aux sections 2 et 3 pour charger les motifs manuellement.

2. Utilisation rapide avec LuaTeX (recommandée)

LuaTeX permet de charger les motifs directement, sans étape de compilation de format préalable. La méthode la plus simple et déjà validée par exécution réelle consiste à copier la liste de motifs (les lignes entre `\patterns{}` et `}` dans `tex/hyph-mg.tex`) directement dans un bloc `\directlua`, comme dans `source/test/sentences_lua.tex` :

```
\directlua{
  local malagasy = lang.new(1)
  malagasy:patterns([[
    .a1li1ka. .a1m2bo1ny. .fa1n2ja1ka1na. [... reste des motifs ...]
  ]])
}
\language1
\lefthyphenmin=2
\righthyphenmin=2
```

Note : une extraction automatique du contenu de `hyph-mg.tex` depuis Lua (via `io.open` et un filtrage du texte) a été tentée pendant le développement de ce guide, mais deux comportements inattendus ont été rencontrés (le caractère `\` écrit littéralement dans un bloc `\directlua` est intercepté par le tokenizer TeX avant d'atteindre Lua). Faute d'avoir pu fiabiliser complètement cette voie dans le temps disponible, la recommandation actuelle reste la copie manuelle ci-dessus, qui est la méthode réellement testée de bout en bout.

Piège réel à éviter : ne jamais mettre de commentaire Lua (`--`) à l'intérieur d'un bloc `\directlua` -- cela corrompt silencieusement l'exécution, sans erreur visible. Commenter en TeX (%) juste avant ou après le

3. Utilisation avec pdfTeX (format INITEX)

pdfTeX exige que `\patterns{}` soit chargé en mode INITEX (génération de format), pas dans une compilation normale. Le malgache doit être isolé sur son propre `\language` (jamais `\language0`, déjà utilisé par l'anglais de `plain.tex`) :

```
tex -ini -interaction=nonstopmode -jobname=hyphmg hyphmg-test.ini
TEXFORMATS=.tex -fmt=hyphmg -interaction=nonstopmode monfichier.tex
```

Le fichier `tex/hyphmg-test.ini` est généré automatiquement par `generate_patterns.py` (voir section 4) et contient déjà la logique d'isolation de langue nécessaire.

4. Régénérer les motifs à partir du corpus

Si vous modifiez `source/malagasy-hyphenation.txt` (le corpus annoté) ou `source/manual-overrides.txt`, régénérez les fichiers livres puis validez :

```
python3 source/generate_patterns.py
# -> tex/hyph-mg.tex, tex/hyphmg-test.ini
python3 source/validate.py
# vérifie clusters, synchronisation, couverture
```

`validate.py` échoue (code de sortie non nul) si un motif scinde un cluster protégé, si `tex/hyph-mg.tex` est désynchronisé du corpus, ou si un override manuel contredit un motif régénéré. Toujours lancer cette commande avant de committer.

5. Exécuter les tests

- `source/test/words.tex` -- césure mot par mot via `\showhyphens` (pdfTeX), comparée à `source/test/expected/words.txt`.
- `source/test/sentences_lua.tex` -- césure en contexte de paragraphe via un callback LuaTeX, comparée à `source/test/expected/sentences.txt`. À préférer à l'approche pdfTeX pure, qui rate les lignes de badness finie (voir documentation, section 3).

```
python3 source/extract_hyphenation.py motfichier.log sortie.txt --mode lua
# ou --mode showhyphens selon le test
```

6. Contribuer une correction ou un nouveau mot

Pour proposer une correction ponctuelle sans toucher au corpus principal, ajoutez le motif à `source/manual-overrides.txt` avec une référence de suivi :

```
mo1n2mot.    # correction du découpage de "monmot" (issue #12)
```

Pour ajouter un mot au corpus principal, l'annoter en notation de Liang dans `source/malagasy-hyphenation.txt` (voir les exemples déjà présents), puis relancer génération et validation (section 4).

Tout ajout de mot beneficie d'une relecture par un locuteur natif avant integration definitive -- contactez l'auteur a l'adresse ci-dessous pour obtenir ou proposer un document de relecture au format identique a celui deja utilise pour les 29 mots actuels.

7. Limites connues (resume)

- Le corpus ne couvre que 29 mots : les mots absents restent simplement non hyphenés (comportement de repli sur, pas une erreur).
- Aucune generalisation a des mots non vus par construction (voir documentation, section 4) -- chaque motif est ancre au mot entier.
- Validation par l'Academia Malgache et integration officielle a hyph-utf8 non encore effectuees.

Pour le detail complet (methodologie linguistique, resultats d'execution reelle, pieges TeX/LuaTeX rencontres), voir *hyph-malagasy-documentation.pdf*.