

Traitement morphologique des formes verbales dérivées du kabyle à l'aide du logiciel NooJ.

Hamid Annouz
Doctorant, Lacnad, Inalco

Introduction

Le traitement automatique des langues naturelles consiste à regrouper la linguistique et la technologie dans le but de décrire et de comprendre l'ensemble des phrases d'une langue donnée à l'aide des programmes de traitement. Mais pour cela, notre outil doit reconnaître les différentes unités de la langue à traiter et les différentes règles par lesquelles ces unités se combinent et forment des phrases.

Le logiciel NooJ est créé par Max Silberztein en 2003 pour dépasser les limites du programme INTEX dont il est issu. Son but est *« de décrire avec exhaustivité et une précision absolue l'ensemble des phrases d'une langue susceptible d'apparaître dans les textes écrit. Ce projet répond à deux besoins : il s'agit d'une part de fournir aux linguistes des outils pour les aider à décrire les langues,, et d'autres part de construire des logiciel capables de traiter automatiquement des textes rédigés en langue naturelle...»* (Silberztein : 2015, 15).

Nous l'avons donc choisi pour un certain nombre de critères qui lui donnent une capacité et un avantage par rapport aux autres outils proposés jusque-là. Parmi les plus importants analyseurs conçus pour le TALN nous citons, *Combinatorial Categorical Grammars (CCG)*, *Head Grammars (HG)*, *Linear Indexed Grammars (LIG)*, *Head-Driven Phrase Structure Grammar (HPSG)*. L'inconvénient de ces outils, contrairement à NooJ, réside dans leur incapacité à

traiter l'ensemble des phénomènes de la langue. Chaque logiciel est conçu pour traiter un phénomène particulier de la linguistique¹.

L'approche proposée par M. Silberztein offre aux linguistes un cadre théorique et méthodologique qui permet le traitement et la description précise et exhaustive de toutes les phrases censées apparaître dans un texte. « *C'est parce que tout projet de description des langues naturelles implique nécessairement la formalisation de nombreux phénomènes linguistiques de nature très diverses que j'ai construit la plateforme NooJ* » (2015, p. 180). Ainsi pour chaque phénomène à traiter, le linguiste dispose d'un outil spécialement conçu.

Sa démarche est inspirée des grammaires génératives² décrites par N. Chomsky (1957). Il propose dans sa théorie quatre niveaux de grammaires auxquelles correspondent quatre niveaux d'analyse :

1. Les grammaires rationnelles : les règles de ces grammaires exigent que le membre gauche doive avoir un seul symbole auxiliaire. Le membre droit contiendra soit un seul symbole terminal (N => chien), soit le mot vide (N => <E>), soit d'un seul symbole terminal suivi d'un seul symbole auxiliaire (N => chat *ADJECTIF*).
2. Les grammaires hors contexte : les règles de ce type exigent que le membre gauche doive être constitué d'un seul symbole auxiliaire. Le membre droit peut avoir différentes combinaisons de symboles terminaux et de symboles auxiliaires (PHRASE => GN voit GN)
3. Les grammaires contextuelles : dans ces grammaires, on peut ajouter des contextes dans les deux parties, comme le contexte PLURIEL dans l'exemple : PLURIEL PHRASE => PLURIEL GN voient

¹ Voir (Silberztein : 2015) pour plus d'informations sur les différents outils.

² Une grammaire générative est un langage formel capable de générer d'autres langages et structures en passant d'un niveau à un autre. Elle est constituée de deux ensembles (à gauche et à droite) et de règles de réécriture.

4. Les grammaires non-restreintes : il s'agit d'aucune restriction ici, toutes les combinaisons sont permises. Exemple : PLURIEL chien PHRASE => PLURIEL GN voient N.

La singularité de l'approche NooJ, contrairement aux autres formalismes, réside dans sa capacité à permettre l'interconnexion de tous ces niveaux de description.

Dans cet exposé, je présente une partie de mon projet de thèse qui consiste à construire un dictionnaire des flexions verbales du kabyle ainsi que d'autres dictionnaires des autres unités linguistiques à savoir : les noms, les adverbes et les prépositions, etc. Ce sont des dictionnaires électroniques à construire à l'aide du logiciel NooJ³. Le but est de constituer une base de données qui sera exploitée pour la constitution de l'ensemble des règles sur lesquelles repose la combinaison des mots et qui permettra la reconnaissance de la phrase correcte.

La liste des verbes traités dans ce travail et leurs modèles de conjugaison sont prise principalement dans le dictionnaire de J-M Dallet (J.-M. DALLET : 1982) et le manuel de conjugaison de kabyle de K. Naït-Zerrad (K. NAÏT-ZERRAD : 1994).

Dans la liste des verbes dérivés, nous avons écarté certaines formes pour les raisons principales suivantes :

- Le sens est loin de la base : *cucci / mčučči* ou *rnu / nnerni*, etc.
- La base n'est pas attestée : *sburr, ssefqed*, etc.
- Le dérivé n'a qu'une très légère nuance sémantique de base : *nekraf* (être ankylosé, avoir les membres raids après une longue station dans une posture inconmode), dérivé de *kref* (être paralysé, ankylosé), ou encore entre *ḍaε* et *ṇṭaε*.

³ Ce genre de dictionnaire ne doit pas être confondu avec les versions électroniques des dictionnaires d'usage. Nous parlons ici d'un outil de Traitement automatique des langues naturelles qui sert à regrouper les différentes informations liées aux mots (flexion, dérivation, propriétés syntaxiques, sémantiques, etc.), afin de permettre au programme informatique de les identifier.

- Le dérivé est un emprunt sans sa forme primaire : *mechur, mehdud, mesud*.

Nous résumons cet exposé dans les cinq étapes suivantes⁴ :

- Construction des modèles de formes dérivées et leur flexion dans la partie grammaire ;
- Ajouts des formes dérivées devant les lemmes dans la partie dictionnaire ;
- Compilation et résultat ;
- Application sur un texte ;
- Quelques problèmes.

Codes et abréviation :

- <LW> => Pour revenir tout en arrière du mot.
- <RW> => Pour aller tout devant.
- <L> => Pour aller de droite à gauche du mot.
- <R> => Pour aller de gauche à droite du mot.
- => Pour effacer un caractère on commençant de droite à gauche.
- <S> => Pour effacer un caractère on commençant de gauche à droite.
- <E> => Pour garder le mot tel qu'il est.
- <D> => Pour doubler le caractère.
- DRV => forme dérivée
- " " => Pour laisser un vide.
- / => Pour passer de la forme fléchie à l'aspect et à la personne qui lui correspond.
- |=> Pour passer d'une forme à une autre à l'intérieur d'un même paradigme.
- ; => Fin de la flexion d'un paradigme.

⁴ Voir notre présentation dans le numéro 9 de cette revue (REB 9-2013).

- + => Pour passer d'une indication à une autre (féminin singulier => f+s).
 : => permet de faire un lien entre la forme de dérivé et son modèle de flexion.
 IMP => Pour désigner l'impératif.
 IMPERINTENS => Impératif intensif.
 PRET => Prétérit.
 PRETNEG => Prétérit négatif.
 INTEN => Aoriste intensif.
 AOR => Aoriste.
 PARTICIPPRET => Participe positif du prétérit.
 PARTICIPNEGPRET => Participe Négatif du prétérit.
 PARTICIPAOR => Participe de l'aoriste simple.
 PARTICIPAOINTPO => Participe de l'aoriste intensif positif.
 PARTICIPAOINTNEG => Participe de l'aoriste intensif négatif.

1. Construction des modèles dérivationnels et leur flexion

1.1 Construction des modèles

```

# Modèles de dérivés en MY-:
#####
DERVMY1= <LW>my/V;
DERVMY2= <L2><S>a<LW>my/V;

```

Figure 2

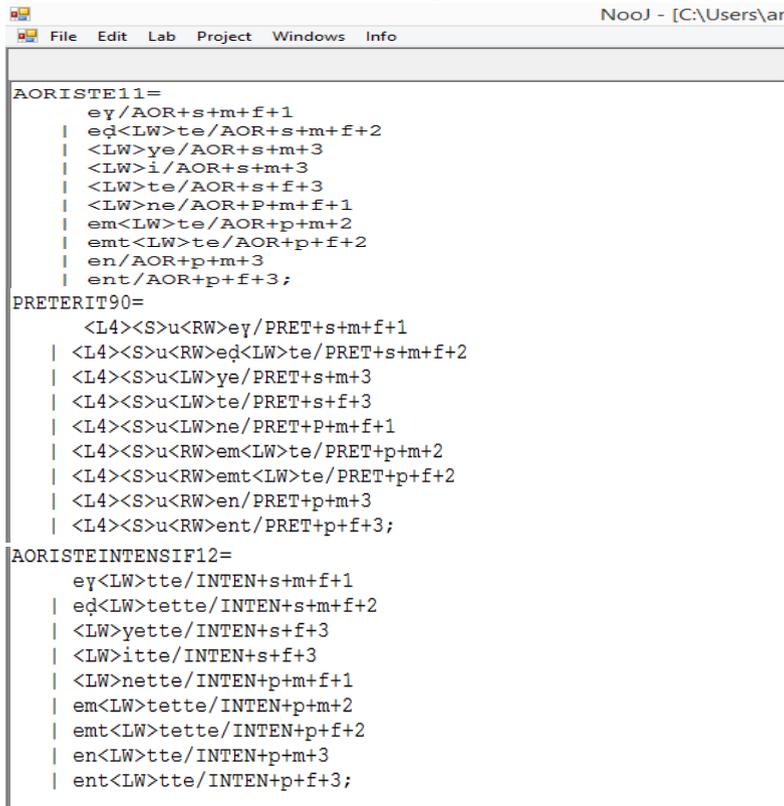
Ces deux exemples nous montrent la procédure suivie pour construire des verbes dérivés à base de leur forme simple. On voit dans la première ligne que pour avoir la forme ‘*myacar*’ il faut juste ajouter ‘*my*’ au début de mot⁵.

⁵ Le code ‘<LW>’ signifie ‘: aller au début du mot’. Pour les autres codes utilisés, voir la liste des codes et symboles.

“DERVMY1” est le nom qu’on a donné à ce modèle qui sera mis devant chaque lemme à qui on ajoute juste le préfixe “my” pour obtenir son dérivé réciproque.

Une fois la forme du dérivé est obtenue nous devons construire son paradigme flexionnel (sa conjugaison complète). Nous procédons par les deux étapes suivantes :

1.2 Construction de leurs paradigmes de flexions



```

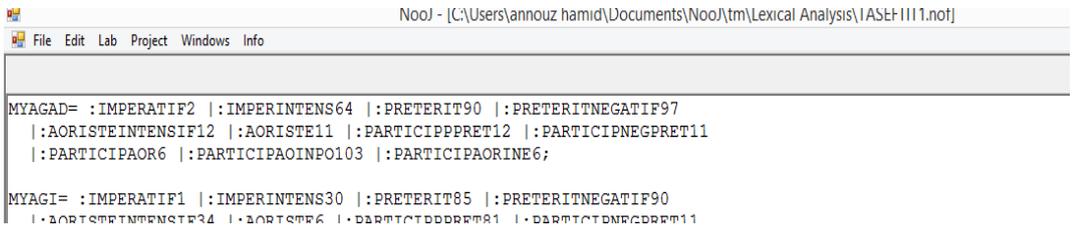
NooJ - [C:\Users\ar
File Edit Lab Project Windows Info
AORISTE11=
  ey/AOR+s+m+f+1
  | eq<LW>te/AOR+s+m+f+2
  | <LW>ye/AOR+s+m+3
  | <LW>i/AOR+s+m+3
  | <LW>te/AOR+s+f+3
  | <LW>ne/AOR+p+m+f+1
  | em<LW>te/AOR+p+m+2
  | emt<LW>te/AOR+p+f+2
  | en/AOR+p+m+3
  | ent/AOR+p+f+3;
PRETERIT90=
  <L4><S>u<RW>ey/PRET+s+m+f+1
  | <L4><S>u<RW>eq<LW>te/PRET+s+m+f+2
  | <L4><S>u<LW>ye/PRET+s+m+3
  | <L4><S>u<LW>te/PRET+s+f+3
  | <L4><S>u<LW>ne/PRET+p+m+f+1
  | <L4><S>u<RW>em<LW>te/PRET+p+m+2
  | <L4><S>u<RW>emt<LW>te/PRET+p+f+2
  | <L4><S>u<RW>en/PRET+p+m+3
  | <L4><S>u<RW>ent/PRET+p+f+3;
AORISTEINTENSIF12=
  ey<LW>tte/INTEN+s+m+f+1
  | eq<LW>tette/INTEN+s+m+f+2
  | <LW>yette/INTEN+s+f+3
  | <LW>itte/INTEN+s+f+3
  | <LW>nette/INTEN+p+m+f+1
  | em<LW>tette/INTEN+p+m+2
  | emt<LW>tette/INTEN+p+f+2
  | en<LW>tte/INTEN+p+m+3
  | ent<LW>tte/INTEN+p+f+3;

```

Figure 3

Il s'agit dans cette étape de construire toute la conjugaison pour tous les verbes. La construction des modèles de flexion séparément de leurs verbes a permis le partage d'un seul paradigme par plusieurs verbes. "AORISTE11" par exemple est partagé par les verbes modèles : *CTIQ*, *ČČAR*, *GGAG*, *LLEXS*, *MBIBB*, *MLIL*, *MMEČČ*, *MMENŸ*, *MMIR*, *MQALAC*, *MQASS*, *MYAF*, *MYENŸ*, *MYIF*, *MEUQQ*, *NNUM*, *QQERŠ*, *QQIM*, *ZWIR*, *SBEDD*, *SBURR*,,) ⁶.

1.3 Rattachement des paradigmes flexionnels à des verbes types⁷



The screenshot shows a window titled 'NooJ - [C:\Users\annouz hamid\Documents\NooJ\fm\Lexical Analysis\IASÉF III 1.nof]'. The interface includes a menu bar with 'File', 'Edit', 'Lab', 'Project', 'Windows', and 'Info'. The main content area displays the following text:

```

MYAGAD= :IMPERATIF2 |:IMPERINTENS64 |:PRETERIT90 |:PRETERITNEGATIF97
|:AORISTEINTENSIF12 |:AORISTE11 |:PARTICIPPPRET12 |:PARTICIPNEGPRET11
|:PARTICIPAOR6 |:PARTICIPAOP103 |:PARTICIPAORINE6;

MYAGI= :IMPERATIF1 |:IMPERINTENS30 |:PRETERIT85 |:PRETERITNEGATIF90
|:PARTICIPINTENSIF34 |:AORISTE6 |:PARTICIPDDPRET11 |:PARTICIPDNEGPRET11

```

Figure 4

Cette figure nous montre le paradigme flexionnel représenté par le modèle "MYAGAD", il est noté devant toutes les entrées de ce modèle.

⁶ Certains résultats qui seront générés par l'application de certaines parties de conjugaison (comme celle de verbe *myagad* à la 1^{ère} personne) sont juste inutiles; elles ne nous poserons aucun problème. Nous avons fait ce choix pour éviter la multiplication des modèles; le paradigme de conjugaison "PRETERIT90" est partagé par d'autres verbes modèles (c'est-à-dire par une centaine de verbes différents).

⁷ Ce sont des verbes intermédiaires qui seront mis devant les entrées de dictionnaires pour générer toutes les flexions.

2. Ajouts des dérivés au dictionnaire

```

#use taseftit.nof

acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD
addi, V+RV+FLX=AGI
adef, V+I+FLX=AFEG
aden, V+T+FLX=AFEG+DRV=DERVS3:SSIWEḌ
aḡen, V+I+FLX=AFEG+DRV=DERVS3:SSIWEḌ+DRV=DERVMS15:MSIREM
ader, V+I+FLX=AFEG+DRV=DERVS3:SSIWEḌ+DRV=DERVMS15:FRIWES+DRV=DERVN6:CRURED
ader, V+T+FLX=AFEG+DRV=DERVTTW2:TTWAREZ
adu, V+T+FLX=ARGU
aeer, V+I+FLX=AFEG
aelay, V+I+FLX=IDYIQ+DRV=DERVS31:SSILWI+DRV=DERVS35:SBERWI
aetad, V+FLX=AKLAL
aeu, V+RV+FLX=ARU+DRV=DERVMY1:MFI+DRV=DERVMS12:MFI
aezag, V+I+FLX=ISHIL
af, V+T+FLX=AF+DRV=DERVTTW2:TTUNEFK+DRV=DERVMY1:MYAF
afeg, V+I+FLX=AFEG+DRV=DERVS3:SSIWEḌ+DRV=DERVSR1:SRIFFEG
afes, V+RV+FLX=AFEG+DRV=DERVTTW2:TTWAREZ
ay, V+RV+FLX=AḠ+DRV=DERVMY1:MYAGAD+DRV=DERVMS12:MQALAC+DRV=DERVS3:SFAD+DRV=DERVTTW1:TTUNEFK
agad, V+I+FLX=AGAD+DRV=DERVS3:SSIWEḌ+DRV=DERVMY1:MYAGAD
agar, V+I+FLX=AGAD+DRV=DERVS3:SSIWEḌ
agat, V+I+FLX=AGAD
agem, V+T+FLX=AFEG

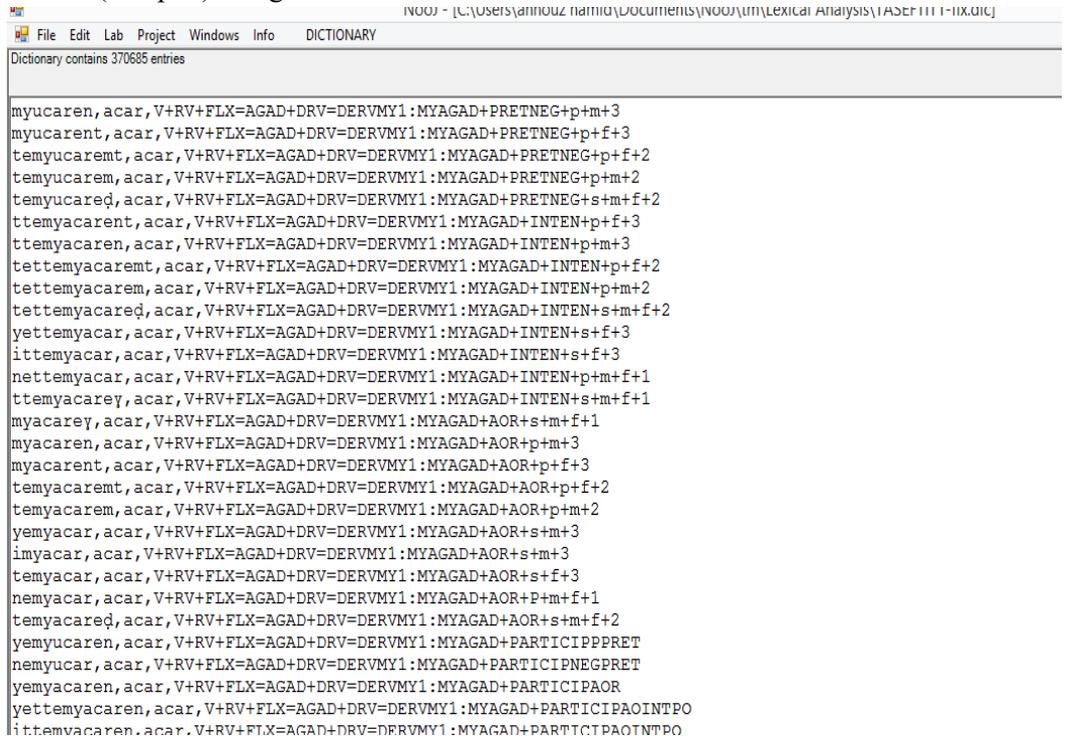
```

Figure 1

On lit dans la première entrée ceci : la forme ‘*acar*’ est un verbe réversible (ou mixte) qui se fléchi selon le modèle ‘AGAD’ et qui a un dérivé dont la forme est celle représentée par le modèle ‘DERVMY1’ (voir *figure 2*), ce dernier se fléchi comme le modèle ‘MYAGAD’ (voir *figure 3* et *figure 4*).

3. Compilation et résultat

Pour avoir toutes les formes fléchies de tous les verbes, le dictionnaire doit être associé (compilé) aux grammaires flexionnelles :



Dictionary contains 370685 entries

```

myucaren, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+PRETNEG+p+m+3
myucarent, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+PRETNEG+p+f+3
temyucarent, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+PRETNEG+p+f+2
temyucarem, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+PRETNEG+p+m+2
temyucaređ, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+PRETNEG+s+m+f+2
ttemyucarent, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+INTEN+p+f+3
ttemyucaren, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+INTEN+p+m+3
tettetmyucarent, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+INTEN+p+f+2
tettetmyucarem, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+INTEN+p+m+2
tettetmyucaređ, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+INTEN+s+m+f+2
yettetmyucar, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+INTEN+s+f+3
ittemyucar, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+INTEN+s+f+3
nettemyucar, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+INTEN+p+m+f+1
ttemyucarey, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+INTEN+s+m+f+1
myacarey, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+AOR+s+m+f+1
myacaren, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+AOR+p+m+3
myacarent, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+AOR+p+f+3
temyucaremt, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+AOR+p+f+2
temyucarem, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+AOR+p+m+2
yemyucar, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+AOR+s+m+3
imyucar, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+AOR+s+m+3
temyucar, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+AOR+s+f+3
nemyucar, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+AOR+p+m+f+1
temyucaređ, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+AOR+s+m+f+2
yemyucaren, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+PARTICIPPPRET
nemyucar, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+PARTICIPNEGPRET
yemyucaren, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+PARTICIPAOR
yettetmyucaren, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+PARTICIPAOINTPO
ittetmyucaren, acar, V+RV+FLX=AGAD+DRV=DERVMY1:MYAGAD+PARTICTPAOINTPO

```

Figure 5

On traduit la première ligne par la phrase suivante : *myucaren* est issu de verbe *acar*, qui se fléchi selon le modèle *AGAD*, ce dérivé qui se réfère au modèle “DERVMY1” et qui se fléchi selon le modèle flexionnel “MYAGAD” est conjugué à la troisième personne du pluriel du prétérit.

La grammaire est constituée de :

- 09 modèles d'impératifs simples.
- 184 modèles d'impératifs intensifs.
- 97 modèles s de prétérit simples.
- 104 modèles de prétérits négatifs.
- 200 modèles d'aoristes intensifs.
- 84 modèles de participes de prétérits positifs.
- 53 modèles de participes de prétérits négatifs.
- 25 modèles de participes d'aoristes simples.
- 196 modèles de participes d'aoristes intensifs positifs.
- 186 modèles de participes d'aoristes intensifs positifs.
- 440 verbes types.
- 3587 lemmes.
- 370685 formes générées.

4. Application sur un texte

Pour éviter quelques ambiguïtés, nous accompagnons l'application du dictionnaire des verbes sur le texte par l'intégration de graphes syntaxiques et de deux autres dictionnaires⁸ : un dictionnaire des affixes de verbes pour épargner quelques affixes homographes de verbes comme : *'-as'* : affixe du verbe régime indirecte à la 3^{ème} p.s.m.f, et *as* : forme simple du verbe (venir) ou *'-yasen'* affixe du verbe régime indirecte à la 3^{ème} p.p.m et le verbe (venir) conjugué au participe de prétérit. Et un dictionnaire de prépositions pour éliminer des formes telles que : *'asen'* dans *'fell-asen'* ou *'fell-am'* qui sont respectivement homographes des verbes : *'as'* (venir) et *am* 'commencer / entreprendre'.

⁸ Voir cette démonstration dans le numéro 9 de cette revue (REB-Volume9-2013).

Beleid At Ali

Tamacalut n Uwayezniw

Macahn... **Rebbi** a tt-**issellu**... A tt-**igerrez am** usaru l... yef yiwien sselṭan, sselṭan ala **Rebbi** yur-s taneṭṭut-is d mmi-is. Netta **yelkem** deg yiwet n tmurt anect-**ila**-tu. Tamesiwt, **iḥer**-ak lḥal ulac win **ikesben am** netta : ama d tferkiwin, ama d ixxamen, ama d idrimen, ama lḥasun d **ayen tebyid**.

Yiwen wass deg wissan n **Rebbi iwala**, dayen, **yeqsed** yakk: wexxam-is, lḥekma-s **tgerrez**, ur **yessi** ara s-**iḥebber** ; lewzir-is **yelha**, **yessen** akk lecyal akken ad ten-**ixdem**, **yettkel** fell-as. Ass-n, akken dimi d-**yewwed** s axxam tameddit, **yettef** mmi-s, **yenna**-yas :

Tura a mmi, kečč aql-**ak tewwdeq** d argaz, **zemrey** ad **ttekley** fell-ak. Ur **yelli** wayen i k-**ixuṣṣen** la di tmusi wala di leeqel. Nekkini **tura** aql-i d amyar, ur **teziq** i **ixeddem** **Rebbi** : **lmut** yef **medden** irkelli. Ili **menmay** a d-**slley** iysan-iw deg wexxam n **Rebbi** : aql-i ad **ruḥey yer** lḥiḡ, a k-**ḡḡey** d kečč i d amseuwweg, d kečč i d lmuḍebber deg wexxam. Yemma-k d taneṭṭut : d kečč ara **iḥarben** fell-as. Keččini a mmi, **ziy**-k d ilemzi, **tewwdeq** i zzwag, lameena **rḡu** ar d-**uyaley** di lḥiḡ, ad ak-**muqem** tameytra, ncalleh.

Yermi **yefka**-yas yakk tisura n yexxamen, n lemẓazen n yedrimen, n nneema, n lekwayeḍ yakk n tferkiwin d trika-**nsem** akken **tella** ; **yermi iweṣṣi**-t akken **ilaq** yef kra n wayen **yellan**. Akken dimi t-**issefhem** akken **iwumen**, **yenna**-yas Seedi mmi-s : A k-**ieuzz** **Rebbi** a baba, akken iji-**teuzzed** imi **tettekled** fell-i, **teḥkiq**-iji rray. Day netta ula d **nekk** aql-i **kksay**-ak aybel : ur **teseqid** ara **ṭhebbred** : **ruḥ** a d-**ṭhuḡeḍ** yef yiman-ik : **kkes** aybel i wul-ik.

Sselṭan **ithenna**. Di syina **yerna** yumayen **ney** data ala i uheggi n weeswin d zzwayel i ttrekba... d wayen **nniden**. Di lweqt-**nni** **medden leḥḥun** ala yef uḍar **ney** win **yessan**, yef leḥwayec : d aserdun, d agewdiw, **ney** d alyem ; **ḥekkun ilan** wid **irekkben** ula yef kra n wefrux akken ur **yezri ḥedd** isem-is, netta atan anect-**ila**-t, d aelayan akken, iḡarren-is d iyezfanen, lameena ur **yettafeg** ara, d tikdi i **leḥḥu**. Di tmurt n sselṭan i yef la d-**nettawi**, **ahat yella** useggas n tikdi **ger** rrawḥ d tuyalin.

Asmi **issewjed** ayen yakk **iḥwaḡ**, **ikker**-d yiwet tsebjit, **ieebba rruḥ**, **itkel** yef **Rebbi**.

Seedi, akken dimi **iruḥ** baba-s, **iqqim** d yiman-is **yeckem**-it aybel n wexxam, n sssuq, n tteḡbir. Lḥasun **yuyal** d leaqel akken **ilzem** : leatba ara **yefk** i fassen-is ur t-**yessi** : ama d ayellet n tferkiwin, ama d ccyel **nniden** yakk, **yessa**-yasen ifellaḥen, aḥal ! Netta, d acu kan, **issefqaḍ** yer wayla-s, **itadaded** i ifellaḥen d inestujar-is, **ney**, mi ara d-**yeḥder** unejmae imeqqraneṣ yef kra **ney** kra, **yettaṭṭaf** amkan n baba-s.

Deg waggur **ney** deg cehrayen imezwira segmi **ruḥ** baba-s **yedḥa**-d rray-is **yewqem**, **yeshleḥ** almi d ulamek. Netta leqraya **yeyra**, kḥaṭer-is **yewsee**, **yeččur** d leḥhama. Ad **termuḍ**, seg wasmi d-**yekker**, **yettwarebba** ala **ger** imnusawen d wuḥdiqen ; **yennufa**-d **yecba**-ten. Mmi-s n sselṭan ara **yessun** ala tigi, berka-t. Seedi, **yerna**-yas **Rebbi** tin ur d-**ssasayan** la idrimen la ttrebga : **yefka**-yas **Rebbi** ṣṣifa d zzyen d sser. **Nekk** a d-**inuy** araw n sselṭan **ideq**-iten-id **Rebbi** seg wakal **yezzeḥ yer** ttef ala i nitni, **yerna** d lḥeqq, akken i **ilaq**. **Nekk** n win iji-**icban**, **yettijid**-iji lḥal mi ara **afey** d derya n lmuḥuk **icement**. M'ara **tettwawdem** lḥaḡa, awfan ad **tennekmal**. Win ara **yasmen** di lḥeqq iwumi ara tent-id-**yeddukkel** **Rebbi**, ur **yumin** ara yis-s. **Yerna**, taneggarut, **ruḥ yelhan**, **yeshfan**, **zeddigen**, ur **ilaq** ara a **tezdey** amkan ur **nemsin**, Seedi, **ilaq** d taneṭṭut ara d-**yinin** fell-as :

Yuli wass, sbahḥur,

A leeyun n ttiḥ :

Tin k-**yezran** d leyniya...

Figure 6

Le texte nous montre que, par l'intégration des différentes formes dérivées, aucune forme verbale n'est épargnée.

5. Quelques problèmes

En prenant un échantillon annoté des unités prises en compte dans ce texte (figure 7), nous constatons déjà plusieurs ambiguïtés :

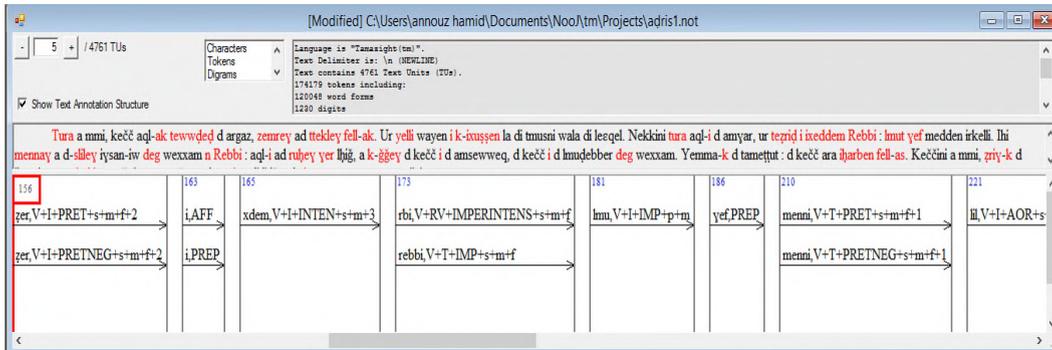


Figure 7

Nous avons pris comme exemple un passage où plusieurs formes sont considérées. Il s'agit d'une réponse à la requête (<V> + <PREP> + <AFF>) c'est-à-dire tous les verbes, affixes et prépositions (les mots en couleur rouge). En bas de texte, on voit la représentation syntaxique d'une phrase et on constate à l'aide des annotations tous les types d'ambiguïtés possibles. Cet exemple nous en montre deux types :

1. Amalgame des valeurs aspectuelles des formes verbales identiques, entre :
 - le prétérite simple et le prétérit négatif dans le cas des verbes *tezırd* et *mennay* et entre la forme de l'impératif intensif du verbe *rbi* (surfaire) et celle de l'impératif simple du verbe *rebbi* (éduquer).
2. Amalgame entre deux unités de deux catégories différentes, à savoir :
 - la préposition *i* (à, pour) et le pronom affixe de la première personne du singulier *i*,
 - *rebbi* (dieu) et la forme simple et intensive de, respectivement, *rebbi* (éduquer) et *rbi* (surfaire),
 - *İmut* (la mort) et la forme de l'impératif simple du pluriel du verbe *İmu* (mariner).

Conclusion

Nous ne sommes qu'au début du travail sur la syntaxe, mais nous pouvons déjà affirmer que la reconnaissance de la phrase kabyle par le logiciel NooJ est possible,

mais difficile. Cette expérience nous a montré d'une manière claire un nombre important de problèmes auxquels nous ne nous sommes pas attendu comme le nombre très important de cas d'homonymies et d'homographies.

Nous nous sommes rendu compte aussi que d'autres problèmes syntaxico-sémantiques nécessiteront un peu plus d'attention et de rigueur dans leur traitement comme dans l'exemple : "*tura isem-is di tewriqt*" qu'on peut interpréter par : (elle a écrit son nom sur une feuille / maintenant son nom est sur la feuille).

Un tel projet nécessitera donc l'implication et l'association de tous les domaines de la linguistique à savoir :

- Le domaine lexical et morphologique : afin de permettre la reconnaissance des unités et la manière dont elles sont construites (flexion, dérivation et composition).

- Le domaine syntaxique : décrire la manière dont les unités se combinent en phrases.

- Le domaine sémantique : décrire la manière dont les sens se combinent pour former le sens global de la phrase.

- Le domaine de la pragmatique : inclure les connaissances générales sur les concepts du monde et celles relatives à la situation de l'énonciation.

Nous signalons enfin le problème du corpus sur lequel tout notre travail sera basé ; il est encore difficile de trouver un texte à exploiter sans vérifier la notation usuelle choisie. Le respect des conventions par les praticiens nous permettra l'utilisation d'un corpus plus large et plus représentatif de la langue.

Bibliographie

FUCHS C., *Linguistique et traitement automatique des langues*, Hachette, France, 1993.

CHAKER S., *Un parler berbère d'Algérie (Kabylie), syntaxe*, Université de Provence, 1983.

- CHAKER S., *Le système dérivationnel verbal berbère, (dialecte kabyle), tome I et II*, thèse de doctorat de troisième cycle, sous la direction de Lionel Galand, soutenu à l'Université de PARIS V, Octobre 1973.
- DALLET J.-M., *Dictionnaire Kabyle-Français*, SELAF, Paris 1982
- DALLET J.-M., *Dictionnaire Français-Kabyle*, SELAF, Paris 1985
- ANNOUZ H., *Construction d'un dictionnaire électronique des flexions verbales du kabyle à l'aide du logiciel NOOJ (Verbes primaires)*, Mémoire pour l'obtention du diplôme de Master 2, sous la direction de K. Naït-Zerrad et de M. Silberztein, soutenu à l'INALCO, Juin 2013.
- CADI K., *Transitivité et diathèse en tarifit, analyse de quelques relations de dépendances lexicales et syntaxique*, Thèse de Doctorat d'Etat, soutenue à l'Université de La sordonne Nouvelle, Paris III, en 1989-1990, sous la direction du professeur David Cohen.
- SILBERZTEIN M., *La formalisation des langues, l'approche NooJ*, ISTE Editions, London, 2015.
- NAÏT-ZERRAD K., *Manuel de conjugaison kabyle*, L'Harmattan, Paris, 1994.
- NAÏT-ZERRAD K., *Grammaire moderne du kabyle, tajerrumt n teqbaylit*, Karthala, Paris, 2001.
- NAÏT-ZERRAD K., *Mémento grammatical et orthographique de berbère, Kabyle-Chleuh-Rifain*, L'Harmattan, Paris, 2011.
- Site internet :
<http://www.nooj5nlp.net>