

La grammaire de la préposition « *d* » avec le logiciel Nooj

Kaci Ferroudja
Lacnad - Inalco, Paris

Introduction

Dans cet article, nous allons traiter la préposition *d* du kabyle (traduite souvent par « et » ou « avec » en français), un parler berbère de Kabylie (Algérie) avec le logiciel Nooj. Pour cela, nous avons utilisé un corpus de Belaid Ait Ali¹ (l'un des premiers écrivains en langue kabyle).

Avec le logiciel Nooj, nous allons essayer de la caractériser par rapport à deux autres morphèmes que sont la particule prédicative (PP) et la particule proximale (PROX) afin d'établir une grammaire de la préposition *d*.

Quelques éléments sur le nom et le morphème *d*

1- Le nom

Le nom est un mot variable en :

a- genre

Le masculin commence en général par la voyelle initiale *a*, *i* ou *u*

axxam « maison » / *imi* « bouche » / *uzzal* « fer »

Le nom féminin se forme sur le masculin par la préfixation et la suffixation de *t*

aqcic « garçon » / *taqcict* « fille »

¹ Dallet J,-M, et Degezelle J-L, 1964, *Les cahiers de Belaid ou la kabylie d'antan*, I, Textes, FDB, Fort-National-National, Algérie.

Le féminin peut aussi dériver d'une racine différente du masculin ex:

argaz « homme » / *tameṭṭut* « femme »

b- nombre

Le pluriel se manifeste par

- le changement de la voyelle initiale et l'ajout d'un suffixe
axxam « maison » / *ixxamen* « maisons »
taxxamt « chambre » / *tixxamin* « chambres »
- Des changements vocaliques
ayaziḍ « coq » / *iyuzaḍ* « coqs »
- un suffixe et des changements vocaliques
izi « mouche » / *izan* « mouches »

c- état

Le noms possède deux états : l'état libre et l'état d'annexion. Quand le nom suit une préposition ou bien quand le nom explicitant le sujet suit le verbe, il est à l'état d'annexion et subit des modifications morphologiques affectant la voyelle initiale : **le maintien** ou **la chute** de la voyelle initiale et l'ajout d'un *w* ou d'un *y* pour les noms masculins.

| | état libre (EL) | état d'annexion (EA) |
|-----------|-----------------|------------------------------------|
| <i>m.</i> | <i>axxam</i> | w <i>exxam</i> « maison » |
| <i>f.</i> | <i>taxxamt</i> | t <i>exxamt</i> « chambre » |

tibḥirt *d* *wexxam*
jardin.EL PREP maison.EA
un jardin et une maison

L'état d'annexion n'est cependant pas marqué pour un certain nombre de noms :

gma « mon frère » (EA = EL) / *tala* « fontaine » (EA = EL)

2- Le morphème *d*

Le morphème *d* peut être :

1- la particule prédicative

Elle assure la prédication d'éléments nominaux : nom, adjectif, substitut indépendant..., traduite en français par (le verbe « être » ou « c'est, ce sont »). Son régime est toujours à l'état libre. Selon Chaker, cette particule avec son régime, présente plus de 10 pour cent des syntagmes prédicatifs en kabyle (Chaker 1983 : 391).

2- la particule proximale

Rattachée au verbe, elle oriente en général le procès vers le locuteur (= vers ici).

3- la préposition

Son régime est un nom à l'état d'annexion.

Le logiciel Nooj

Le logiciel Nooj est une plate-forme de développement linguistique. Il permet la description des langues naturelles ou techniques. Un logiciel créé par Max Silberstein en 2003 (en améliorant les fonctionnalités d'INTEX), il permet de présenter et de formaliser les langues naturelles dans le but de les appliquer en temps réel à de gros corpus. 20 modules de langues différentes sont déjà disponibles et nous espérons ajouter un module pour le kabyle. Nooj nous fournit des outils pour décrire les neuf (8) niveaux linguistiques :

- orthographe
- typographie
- morphologie flexionnelle
- morphologie dérivationnelle
- syntaxe locale
- syntaxe structurelle
- syntaxe transformationnelle
- Sémantique

Les étapes de l'analyse

a. Construire un dictionnaire des noms

Nous avons récupéré les noms dans le corpus qui sont à l'état d'annexion et nous avons construit notre dictionnaire. L'entrée lexicale doit être à l'état libre et au singulier.

```

NooJ - [dic,noms.dic]
File Edit Lab Project Windows
Dictionary contains 504 entries

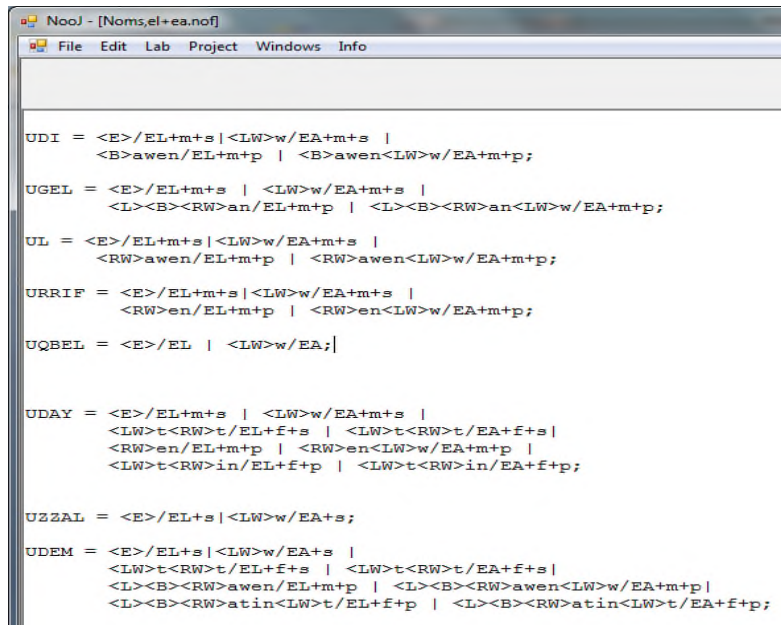
# Special Characters: '\''
#use Noms,el+ea.nof

udi, N+FLX=UDI
ugel, N+FLX=UGEL
ul, N+FLX=UL
urrif, N+FLX=URRIF
uqbel, ADV+FLX=UQBEL
ulac, ADV+FLX=UQBEL
uday, N+FLX=UDAY
ucbiḥ, N+FLX=UDAY
usu, N+FLX=UDI
ucmit, N+FLX=UDAY
uḥdiq, N+FLX=UDAY
uzzal, N+FLX=UZZAL
udem, N+FLX=UDEM
  
```

Un mot comme *udi* « beurre » est défini dans le dictionnaire comme un nom (N) qui va se fléchir (FLX) dans la grammaire sous le nom ou le code (On peut mettre aussi un code, n'importe lequel, par exemple L1) de « *UDI* » (les entrées lexicales dans la grammaire seront toujours en majuscules à la différence du dictionnaire, qui seront en minuscule).

b. Construire la grammaire

Chaque mot dans le dictionnaire doit avoir sa flexion dans la grammaire. On aura quatre (4) modèles de flexions multipliés par deux avec l'état d'annexion pour chacun d'entre eux (huit formes). Chaque nom aura son singulier et son pluriel, son féminin et son masculin sauf pour les adverbes, interrogatifs, préposition et noms d'action verbale qui ne partagent avec le nom que l'une de ces modalités qui est l'état. Il peut avoir aussi des noms qui n'ont pas l'une ou l'autre des modalités du genre ou du nombre.



```

NooJ - [Noms,el+ea.nof]
File Edit Lab Project Windows Info

UDI = <E>/EL+m+s | <LW>w/EA+m+s |
      <B>awen/EL+m+p | <B>awen<LW>w/EA+m+p;

UGEL = <E>/EL+m+s | <LW>w/EA+m+s |
       <L><B><RW>an/EL+m+p | <L><B><RW>an<LW>w/EA+m+p;

UL = <E>/EL+m+s | <LW>w/EA+m+s |
     <RW>awen/EL+m+p | <RW>awen<LW>w/EA+m+p;

URRIF = <E>/EL+m+s | <LW>w/EA+m+s |
       <RW>en/EL+m+p | <RW>en<LW>w/EA+m+p;

UQBEL = <E>/EL | <LW>w/EA;|

UDAY = <E>/EL+m+s | <LW>w/EA+m+s |
       <LW>t<RW>t/EL+f+s | <LW>t<RW>t/EA+f+s |
       <RW>en/EL+m+p | <RW>en<LW>w/EA+m+p |
       <LW>t<RW>in/EL+f+p | <LW>t<RW>in/EA+f+p;

UZZAL = <E>/EL+s | <LW>w/EA+s;

UDEM = <E>/EL+s | <LW>w/EA+s |
       <LW>t<RW>t/EL+f+s | <LW>t<RW>t/EA+f+s |
       <L><B><RW>awen/EL+m+p | <L><B><RW>awen<LW>w/EA+m+p |
       <L><B><RW>atin<LW>t/EL+f+p | <L><B><RW>atin<LW>t/EA+f+p;
  
```

Par exemple, le nom *udi* a le code UDI dans la grammaire : l'opérateur <E> indique l'état libre au masculin singulier (EL+m+s). Un déplacement au début du mot et l'ajout d'un *w* (<LW>w) met le mot à l'état d'annexion (EA+m+s).

c. Compilation du dictionnaire

Après avoir terminé avec la grammaire, il faut l'associer au dictionnaire des noms. L'opération de compilation et de flexion génère automatiquement toutes les formes possibles pour chaque entrée lexicale du dictionnaire grâce à la grammaire formalisée.

```
#use Noms,el+ea.nof
#
udi,udi,N+FLX=UDI+EL+m+s
udawen,udi,N+FLX=UDI+EL+m+p
wudawen,udi,N+FLX=UDI+EA+m+p
wudi,udi,N+FLX=UDI+EA+m+s
ugel,ugel,N+FLX=UGEL+EL+m+s
uglan,ugel,N+FLX=UGEL+EL+m+p
wuglan,ugel,N+FLX=UGEL+EA+m+p
wugel,ugel,N+FLX=UGEL+EA+m+s
ul,ul,N+FLX=UL+EL+m+s
ulawen,ul,N+FLX=UL+EL+m+p
wulawen,ul,N+FLX=UL+EA+m+p
wul,ul,N+FLX=UL+EA+m+s
```

La figure ci-dessus, est le dictionnaire des flexions. Il rend compte de toutes les formes possibles que peut avoir une entrée lexicale du dictionnaire des noms. Ainsi *udi* « beurre », on aura les 4 formes possibles dans le kabyle

Udi/ Udawen/Wudawen/Wudi.

S'il y a une forme générée qui ne correspond pas à ce qui existe dans la langue, il faut vérifier d'où provient l'erreur. Si l'erreur est dans la grammaire, il faut la corriger sinon voir le dictionnaire. Après chaque rectification, c'est primordial d'enregistrer et de compiler le dictionnaire encore une fois pour que Nooj utilise les dernières modifications.

Application du dictionnaire :

Après avoir lancé la requête permettant de localiser les séquences *d+ Nom à l'état d'annexion*, on obtient un tableau de concordances :

The screenshot shows two windows from the NooJ software. The left window, titled 'belcid at eli.not [Modified]', is used for defining search patterns. It shows the 'Pattern is:' section with 'a NooJ regular expression:' selected, and the pattern 'd<N+EA>' entered. The right window, titled 'Concordance for Text belcid at eli.not [Modified]', displays the search results in a table format. The table has four columns: 'Text', 'Before', 'Seq.', and 'After'. The search results are as follows:

| Text | Before | Seq. | After |
|--|--------|------------|--------------------|
| <i>_Ruħemt a tetħsem... Netta teqqim</i> | | d wemyar | -is. Sakin, teħku- |
| <i>d lbeƣd am yemma-s,</i> | | d yesselma | -s, tamusni tesser |
| <i>yerr-d yer utemmu, yaf-</i> | | d welleħ | Seedi, ala amkar |
| <i>ma seđdan-d syagi wergaz</i> | | d tmeħħut | rekben asewdiw |
| <i>azekka ! Win yeqqar-as « argaz</i> | | d tmeħħut | !, wayeđ yin'as « |
| <i>k-steqsiy : ur seđdan ara</i> | | d webrid | -agi wergaz t-me |
| <i>m'ur twalađ ara argaz</i> | | d tmeħħut | ... Win yin'as aki |
| <i>a nzer... Ur as-qqaren</i> | | d teryel | ... Almi d teryel. |
| <i>as-qqaren d teryel,... Almi</i> | | d teryel | . Ula d bu lebraq, |
| <i>ten-yeskad, yettiwhim di teqciť</i> | | d weqciť | : ma d ameksa, y |
| <i>di zzman n zik, netta</i> | | d tmeħħut | -is, d sebea warr: |
| <i>ten-yessemlal Sidi Rebbi, wergaz</i> | | d tmeħħut | , seeħun-d ala arr |

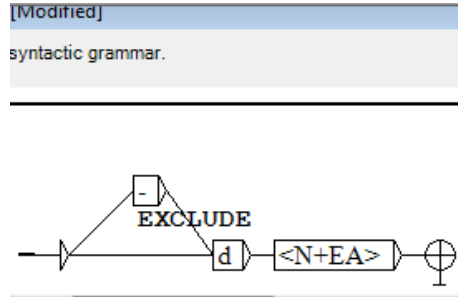
The 'Query' field at the bottom of the concordance window shows 'd' and the result count is '246/246'.

Hormis la préposition « *d* » suivie des noms à l'état d'annexion, on trouve également des particules proximales suivies de noms à l'état d'annexion et des particules prédicatives suivies de noms qui ont l'apparence de l'état d'annexion.

Les problèmes rencontrés sont donc l'homographie des 3 particules mentionnées plus haut et les noms dont l'état d'annexion n'est pas marqué.

Désambiguïsation

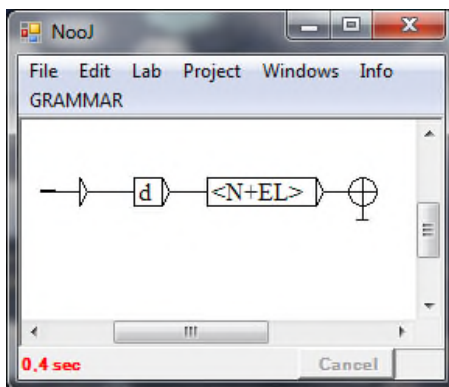
Il faut essayer de résoudre les problèmes rencontrés. Pour ce qui est de la particule prédicative et la particule proximale, des grammaires syntaxiques permettent à NooJ de distinguer entre les 3 morphèmes.



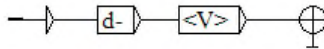
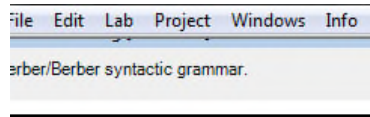
La grammaire ci-dessus concerne la préposition : la particule proximale étant liée au verbe par un tiret dans l'orthographe usuelle, si on exclut la séquence *-d* ou *d-*, l'ambiguïté est levée. Dans l'exemple suivant, Nooj reconnaîtra donc que *-d* n'est pas une préposition mais une particule proximale :

| | |
|-------------------------|---------------|
| <i>yufa-d</i> | <i>weqic.</i> |
| SUJms.trouver.PRET-PROX | garçon |
| Le garçon a trouvé | |

Le graphe ci-dessous représente la particule prédicative, Nooj la reconnaîtra car elle est suivie d'un nom à l'état libre.



La dernière grammaire ci-dessous, est consacré à la particule proximale placée avant le verbe.



Conclusion

Cette désambiguïisation est bien entendu très partielle, car le plus gros du travail reste à faire. Il faut élaborer un dictionnaire des verbes et un autre graphe qui rend compte des différents affixes compléments qui peuvent se glisser entre la particule proximale et le verbe et rassembler tous les graphes dans le même graphe pour une grammaire de *d*. Pour les noms invariables en état, il faudra créer des grammaires pour désambiguïser la majorité des cas. Pour les cas irréductibles, le passage à un autre niveau de l'analyse - le niveau sémantique - permettra peut-être de les résoudre.

Abréviations et codes de Nooj

1- Abréviations

PREP: préposition.

PRET : prétérit. .

PP: particule prédicative;

Prox : proximale.

EL: état libre.

EA: état d'annexion.

N: nom.

ADV: adverbe.

AV: nom d'action verbale.

ADJ: adjectif.

ITR: interrogatif.

DEM: démonstratif.

f: féminin.

m: masculin.

s: singulier.

p: pluriel.

2- Codes de Nooj

- <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.
- / => 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.
- + => Pour passer d'une indication à une autre (féminin singulier => f+s).

Bibliographie:

- Chaker S., 1983 - *Un parler berbère d'Algérie (Kabylie), Syntaxe*, Publication universitaire de Provence, Paris.
- Dallet J.-M. & Degezelle J.-L., 1964 - *Les cahiers de Belaid ou la Kabylie d'antan*, I, Textes, FDB, Fort-National, Algérie
- Naït-Zerrad K., 2001 - *Grammaire moderne du kabyle*, Karthala, Paris.
- Silberztein M. 2003. NOOJ'manual. <http://www.nooj4nlp.net>