

# DÉVELOPPEMENTS INFORMATIQUES POUR LE TRAITEMENT DU SYRIAQUE SOUS UNITEX

*Laurent Kevers*  
*kevers@tedm.ucl.ac.be*

Les développements présentés ici ont été effectués par l'équipe du CENTAL.

## Table des matières

A) Introduction.....	2
B) Architecture générale du logiciel de lemmatisation.....	2
C) Le syriaque.....	5
1. Difficultés rencontrées et solutions apportées.....	5
Prétraitement du texte.....	5
La recherche via « Locate ».....	5
Le dictionnaire électronique DELA Syriaque.....	5
L'affichage « de droite à gauche ».....	5
L'affichage des caractères.....	6
L'adaptation à l'interface de lemmatisation (UnitexL).....	6
Globalement.....	6
2. Ressources.....	8
Alphabet.....	8
Graphe de découpe en phrases.....	8
Catégories grammaticales.....	9
Le DELA Syriaque.....	9
Rapport de compatibilité.....	9
Contenu du dictionnaire.....	14

## A) INTRODUCTION

Unitex est un logiciel permettant l'analyse lexicale et syntaxique des textes. Les processus d'analyse s'appuient sur des ressources linguistiques externes. Le développement d'une nouvelle langue sous Unitex demande des adaptations informatiques d'une part, et la création des ressources linguistiques adaptées d'autre part. UnitexL est une version d'Unitex modifiée pour répondre aux besoins de l'analyse de sources en langue grecque ancienne. Trois types d'adaptations ont été effectuées:

- l'ajout d'un module de référencement;
- l'ajout d'une interface de lemmatisation;
- la création des ressources linguistiques externes propres au grec ancien.

A cette date, UnitexL est opérationnel pour le grec ancien.

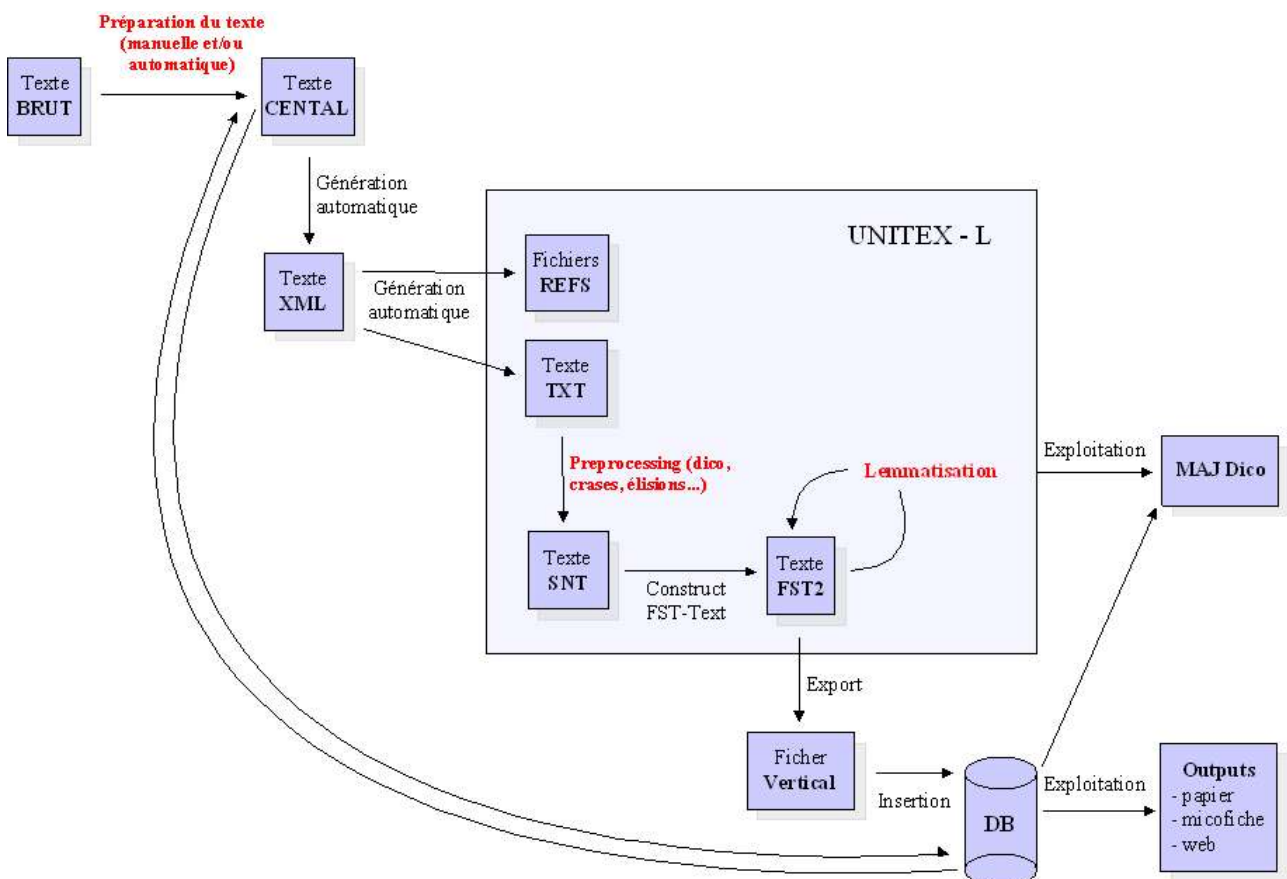
Le travail d'adaptation de ces logiciels au syriaque s'est fait avec l'aide de Sébastien Paumier.

L'adaptation du logiciel au syriaque a été envisagée selon deux phases successives:

- l'adaptation complète du syriaque sous la version classique d'Unitex (les difficultés relèvent tant du domaine informatique que linguistique);
- l'adaptation de l'interface de lemmatisation (dépendant des solutions apportées au point précédent).

La première partie (B) décrit les formats successifs des textes durant les différentes phases du traitement. La seconde partie (C) présente les difficultés rencontrées pour le syriaque et les premières ressources linguistiques disponibles.

## B) ARCHITECTURE GÉNÉRALE DU LOGICIEL DE LEMMATISATION



Lors du processus complet de lemmatisation, le texte passe par différents états, résumés par le schéma ci-dessus. Les étapes indiquées en rouge sont celles où l'intervention du lemmatiseur est prépondérante.

### Etape 1

Le texte de départ est obtenu par saisie manuelle, par scanning avec reconnaissance de caractères ou par récupération d'une version digitale préexistante. Il s'agit d'obtenir le **texte brut** Unicode (UTF-16LE). Il faudra éventuellement adapter le texte depuis un encodage tel que l'ASCII ou le Betacode vers Unicode. Le texte doit être le plus « propre » possible, une série de vérifications s'impose: remplacement de caractères latins par les caractères grecs correspondants, correction d'éventuelles coquilles, toilette typographique (traits d'union, tirets, etc.)...

### Etape 2

Le **texte brut** est converti automatiquement en **texte au format CENTAL**. Il s'agit du continu de référence. Cette version du texte intègre des indications de découpage qui seront utilisées pour la référencement (\$0: oeuvre, \$1->\$7: personnalisable, \$8: page, \$9: ligne \$R: niveaux de rupture des contextes). Elles sont insérées de manière manuelle ou automatique en fonction de la provenance du **texte brut**. Les modifications des références ou du texte doivent être apportées sur ce **texte au format CENTAL**. Les modifications ultérieures imposeront de repasser par cette étape.

### Etape 3

Le **texte au format CENTAL** est converti automatiquement en **texte XML**. Cet état est utilisé afin de créer les fichiers nécessaires au fonctionnement d'UnitexL. En observant l'imbrication des balises, on retrouve le découpage spécifié dans le **texte au format CENTAL**.

**\$0=Marc \$1=1 \$2=t \$9=1 \$R=012**  
 ΚΑΤΑ ΜΑΡΚΟΝ

**\$0=Marc \$1=1 \$2=1 \$9=1 \$R=012**  
 Ἀρχὴ τοῦ εὐαγγελίου Ἰησοῦ Χριστοῦ [υἱοῦ θεοῦ].

**\$0=Marc \$1=1 \$2=2 \$9=1 \$R=012**  
 Καθὼς γέγραπται ἐν τῷ Ἠσαΐα τῷ προφήτῃ,  
 Ἰδοὺ ἀποστέλλω τὸν ἀγγελὸν μου πρὸ προσώπου σου,  
 ὃς κατασκευάσει τὴν ὁδὸν σου·

**\$0=Marc \$1=1 \$2=3 \$9=1 \$R=012**  
 φωνὴ βοῶντος ἐν τῇ ἐρήμῳ,  
 Ἐτοιμάσατε τὴν ὁδὸν κυρίου,  
 εὐθείας ποιεῖτε τὰς τρίβους αὐτοῦ—

**\$0=Marc \$1=1 \$2=4 \$9=1 \$R=012**  
 ἐγένετο Ἰωάννης βαπτίζων ἐν τῇ ἐρήμῳ καὶ κηρῦσσω  
 βάπτισμα μετανοίας εἰς ἄφεσιν ἁμαρτιῶν.

**\$0=Marc \$1=1 \$2=5 \$9=1 \$R=012**  
 καὶ ἐξεπορεύετο  
 πρὸς αὐτὸν πᾶσα ἡ Ἰουδαία χώρα καὶ οἱ Ἰεροσολυμίται  
 πάντες, καὶ ἐβαπτίζοντο ὑπ' αὐτοῦ ἐν τῷ Ἰορδάνῃ  
 ποταμῷ ἐξομολογούμενοι τὰς ἁμαρτίας αὐτῶν.

**\$0=Marc \$1=1 \$2=6 \$9=1 \$R=012**  
 καὶ ἦν  
 ὁ Ἰωάννης ἐνδεδυμένος τρίχας καμήλου καὶ ζῶνῃν  
 δερματίνην περὶ τὴν ὀσφύν αὐτοῦ, καὶ ἐσθίων ἀκρίδας  
 καὶ μέλι ἄγριον.

**\$0=Marc \$1=1 \$2=7 \$9=1 \$R=012**  
 καὶ ἐκήρυσσεν λέγων, Ἔρχεται ὁ ἰσχυρότερός  
 μου ὀπίσω μου, οὗ οὐκ εἰμι ἰκανὸς κύψας  
 λύσαι τὸν ἱμᾶντα τῶν ὑποδημάτων αὐτοῦ·

```
<?xml version="1.0" encoding="UTF-16LE" ?>
<DOC source="C:\TEMP\marcbis.centaf">
  <R0 utxShort="Evangiles" utxBreak="y" utxSelect="y">
    <R1 utxShort="Marc" utxBreak="y" utxSelect="y">
      <R2 utxShorts="1" utxBreak="y" utxSelect="y">
        <R3 utxShort="t">
          <R9 utxShort="1">
            <![CDATA[ ΚΑΤΑ ΜΑΡΚΟΝ.
          ]]>
        </R9>
      </R3>
      <R3 utxShort="1">
        <R9 utxShort="1">
          <![CDATA[ Ἀρχὴ τοῦ εὐαγγελίου Ἰησοῦ Χριστοῦ [υἱοῦ θεοῦ].
          ]]>
        </R9>
      </R3>
      <R3 utxShort="2">
        <R9 utxShort="1">
          <![CDATA[ Καθὼς γέγραπται ἐν τῷ Ἠσαΐα τῷ προφήτῃ,
          ]]>
        </R9>
        <R9 utxShort="2">
          <![CDATA[ Ἰδοὺ ἀποστέλλω τὸν ἀγγελὸν μου πρὸ προσώπου σου,
          ]]>
        </R9>
        <R9 utxShort="3">
          <![CDATA[ ὃς κατασκευάσει τὴν ὁδὸν σου·
          ]]>
        </R9>
      </R3>
      <R3 utxShort="3">
        <R9 utxShort="1">
          <![CDATA[ φωνὴ βοῶντος ἐν τῇ ἐρήμῳ,
          ]]>
        </R9>
        <R9 utxShort="2">
          <![CDATA[ Ἐτοιμάσατε τὴν ὁδὸν κυρίου,
          ]]>
        </R9>
        <R9 utxShort="3">
          <![CDATA[ εὐθείας ποιεῖτε τὰς τρίβους αὐτοῦ —
          ]]>
        </R9>
      </R3>
      <R3 utxShort="4">
        <R9 utxShort="1">
          <![CDATA[ ἐγένετο Ἰωάννης βαπτίζων ἐν τῇ ἐρήμῳ καὶ κηρῦσσω
          ]]>
        </R9>
      </R3>
    </R1>
  </R0>
</DOC>
```

Texte au format CENTAL

Texte XML

#### *Etape 4*

Le texte traité sous UnitexL et ses fichiers associés sont créés automatiquement à partir du **texte XML**. Les diverses ressources linguistiques d'Unitex (alphabet, dictionnaires, graphe de découpe en phrases, autres graphes de prétraitement...) interviennent pour préparer le texte. Le fichier de travail pour la lemmatisation est le **FST-Text**.

#### *Etape 5*

L'interface de lemmatisation permet d'attribuer une et une seule analyse grammaticale pour chaque mot du texte dans son contexte. Une fois celui-ci entièrement traité, on en tire un **fichier d'export**. Ce fichier est confronté au continu de départ pour remplir la base de données qui servira à produire les outputs habituels: concordances et listes sur papier, sur microfiches ou en format digital. L'enrichissement du dictionnaire peut se faire sur la base des listes de nouveaux éléments générées par UnitexL ou en allant rechercher les informations dans la base de données.

## **C) LE SYRIAQUE**

### **1. Difficultés rencontrées et solutions apportées**

#### **Prétraitement du texte**

Lors des premiers essais d'exploitation d'un texte syriaque sous Unitex, l'étape de « tokenization », c.-à-d. de découpe en unités lexicales, n'était pas fonctionnelle. Cette étape étant indispensable, nous avons développé la ressource nécessaire à son exécution: le fichier d'alphabet (Alphabet.txt, voir point 2). Suite à cet ajout, la découpe en tokens a pu être effectuée. Pour compléter le prétraitement, un premier graphe de découpe en phrase a également été réalisé (voir point 2).

#### **La recherche via « Locate »**

Dans un premier temps, il s'est avéré que la recherche de motifs dans le texte ne donnait pas de résultat satisfaisant. Les seuls éléments localisables étaient les lettres simples. Nous avons pu localiser l'origine du problème. La fonction de la bibliothèque Unicode d'Unitex `u_is_a_letter(...)` a été modifiée afin d'intégrer le support du syriaque (ajout de l'appel à la fonction `u_is_syriac_letter(...)`). Cette modification a permis de rendre effective la recherche de formes dans le texte.

#### **Le dictionnaire électronique DELA Syriaque**

Afin de pouvoir profiter de toutes les possibilités de recherche du « Locate », un premier dictionnaire électronique au format DELA (Dictionnaire Electronique du LADL) a été créé. Même s'il s'avère être peu étendu, voire incorrect, cette ressource nous a permis de tester l'application du dictionnaire sur le texte ainsi que la recherche par codes grammaticaux. Ces deux points n'ont pas posé de problème particulier. Dans la perspective de la construction d'un « vrai » premier dictionnaire, des catégories grammaticales ont été définies par A. Schmidt et J.-C. Haelewyck (voir la liste au point 2).

La version du DELA syriaque est bien entendu incomplète et devra être enrichie au fur et à mesure

du traitement des textes. Actuellement, plusieurs lignes du dictionnaire ne sont pas compatibles avec le format du DELA (voir le rapport de compatibilité au point 2) et devront par conséquent être corrigées. Les codes grammaticaux utilisés ne correspondent que partiellement aux codes de la nomenclature car ils ont été définis postérieurement à la construction du DELA. Un travail d'uniformisation entre le dictionnaire et la nomenclature des codes grammaticaux doit donc être effectué.

### **L'affichage « de droite à gauche »**

Une autre difficulté de l'exploitation du syriaque sous Unitex est l'affichage du texte de la droite vers la gauche. De nombreux composants d'Unitex n'étaient pas adaptés pour ce type d'affichage (fenêtre du texte, graphe du texte, concordances, divers champs de saisie...). Avec l'aide de Sébastien Paumier, tous ces éléments ont été modifiés afin de supporter complètement le sens de lecture. A noter que ces améliorations profiteront à tous les langages « de droite à gauche » et donc, entre autres, à l'arabe.

### **L'affichage des caractères**

Une des caractéristiques du syriaque est le comportement graphique de certaines lettres en fonction du contexte dans lequel elles sont placées. En effet, certaines lettres possèdent plusieurs graphies qui sont utilisées pour relier la lettre à telle ou telle autre lettre. Ce comportement n'est malheureusement géré ni par Unicode<sup>1</sup> ni par Java, le langage de programmation dans lequel est écrit Unitex. Les recherches effectuées sur Internet et les diverses personnes contactées (spécialiste Java à l'Université de Marne-la-Vallée, contact avec SIL<sup>2</sup> pour leur système Graphite<sup>3</sup>, contact avec George Kiraz de l'organisation de Beth Mardutho<sup>4</sup>) n'ont pas permis de dégager une solution à ce problème.

### **L'adaptation à l'interface de lemmatisation (UnitexL)**

Indépendamment des points de blocages cités ci-dessus, l'adaptation de l'interface de lemmatisation (UnitexL) au syriaque pose plusieurs problèmes.

Comme expliqué plus haut, les textes doivent passer par différentes phases de prétraitement afin de produire l'ensemble des fichiers nécessaires au fonctionnement d'UnitexL. Il s'agit tout d'abord de créer les ressources utiles pour la création des références. Les programmes opérationnels pour le grec sont, dans ce cas, déjà compatibles pour le syriaque.

L'interface ayant été construite pour le grec ancien sans réelle conscience de la difficulté que pouvait représenter une langue telle que le syriaque (écriture de droite à gauche, transformation de la graphie des lettres, problème des points qui font varier les mots...), il y aura des adaptations à effectuer à divers niveaux. L'évaluation de l'ampleur de ces modifications est cependant assez difficile à réaliser, d'autant plus qu'il est probable que certains obstacles ne nous soient pas encore apparus. Ces adaptations n'ont pas encore été effectuées.

Enfin, nous devons également faire face à une difficulté au niveau de l'organisation du développement. En effet, l'interface de lemmatisation UnitexL a été construite sur la base d'Unitex

---

1 Page principale: <http://www.unicode.org/>

Définition de la plage syriaque: <http://www.unicode.org/charts/PDF/U0700.pdf>

Chapitre sur les « Middle Eastern Scripts »: <http://www.unicode.org/versions/Unicode4.0.0/ch08.pdf>

2 <http://www.sil.org/>

3 [http://scripts.sil.org/cms/scripts/page.php?site\\_id=nrsi&cat\\_id=RenderingGraphite](http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&cat_id=RenderingGraphite)

4 <http://www.bethmardutho.org>









Line 92: unprotected comma in lemma  
ሥዕል፣ገደብ፣ሥዕል፣ገደብ፣.PART  
Line 97: unprotected comma in lemma  
ሥራ፣ገ፣ገ፣.I+ADV  
Line 98: unprotected comma in lemma  
ሥራ፣ገ፣ገ፣.PRO+PERS  
Line 113: unprotected comma in lemma  
፣ገ፣፣.I+PREP  
Line 114: unprotected comma in lemma  
፣ገ፣፣.PRO+DEM  
Line 116: unprotected comma in lemma  
፣፣፣.PART  
Line 118: unprotected comma in lemma  
ሥራ፣፣.PART  
Line 120: unprotected comma in lemma  
ሥራ፣፣.PART  
Line 123: unprotected comma in lemma  
፣፣፣.PART  
Line 127: unprotected comma in lemma  
ሥራ፣፣.I+ADV  
Line 128: unprotected comma in lemma  
ሥራ፣፣.PRO+PERS  
Line 132: unprotected comma in lemma  
ሥራ፣፣.PART  
Line 133: unprotected comma in lemma  
ሥራ፣፣.N+C  
Line 135: unprotected comma in lemma  
ሥራ፣፣.I+PREP  
Line 136: unprotected comma in lemma  
ሥራ፣፣.PRO+PERS  
Line 140: unprotected comma in lemma  
ሥራ፣ሥራ፣፣፣.PART  
Line 141: unprotected comma in lemma  
ሥራ፣ሥራ፣፣፣.N+C  
Line 142: unprotected comma in lemma  
ሥራ፣ሥራ፣፣፣.PRO+PERS  
Line 143: unprotected comma in lemma  
ሥራ፣ሥራ፣፣፣.N+C  
Line 144: unprotected comma in lemma  
ሥራ፣ሥራ፣፣፣.PRO+PERS  
Line 149: unprotected comma in lemma  
ሥራ፣፣፣.I+CONJ  
Line 150: unprotected comma in lemma  
ሥራ፣፣፣.etpe.  
Line 151: unprotected comma in lemma  
ሥራ፣፣፣.N+C  
Line 152: unprotected comma in lemma  
ሥራ፣፣፣.PRO+PERS  
Line 154: unprotected comma in lemma  
ሥራ፣፣፣.PART  
Line 155: unprotected comma in lemma  
ሥራ፣፣፣.N+C  
Line 158: unprotected comma in lemma  
ሥራ፣፣፣.PART  
Line 159: unprotected comma in lemma  
ሥራ፣፣፣.N+C  
Line 163: unprotected comma in lemma  
ሥራ፣ሥራ፣፣፣.N+C  
Line 164: unprotected comma in lemma  
ሥራ፣ሥራ፣፣፣.PRO+PERS  
Line 165: unprotected comma in lemma  
ሥራ፣፣፣.PART  
Line 166: unprotected comma in lemma  
ሥራ፣፣፣.N+C  
Line 170: unprotected comma in lemma  
ሥራ፣ሥራ፣፣፣.N+C  
Line 171: unprotected comma in lemma  
ሥራ፣ሥራ፣፣፣.PRO+PERS  
Line 172: unprotected comma in lemma  
ሥራ፣፣፣.PART  
Line 173: unprotected comma in lemma  
ሥራ፣፣፣.N+C  
Line 183: unprotected comma in lemma  
ሥራ፣ሥራ፣፣፣.N+C  
Line 184: unprotected comma in lemma  
ሥራ፣ሥራ፣፣፣.PRO+PERS  
Line 185: unprotected comma in lemma  
ሥራ፣፣፣.PART  
Line 186: unprotected comma in lemma  
ሥራ፣፣፣.N+C  
Line 187: unprotected comma in lemma  
ሥራ፣፣፣.PART  
Line 188: unprotected comma in lemma  
ሥራ፣፣፣.N+C



Line 265: unprotected comma in lemma  
ሥራገጣጣጣጣ .pa.  
Line 268: unprotected comma in lemma  
ሥራገጣጣጣጣ .I+PREP  
Line 269: unprotected comma in lemma  
ሥራገጣጣጣጣ .PRO+PERS  
Line 270: unprotected comma in lemma  
ሥራገጣጣጣጣ .I+PREP  
Line 271: unprotected comma in lemma  
ሥራገጣጣጣጣ .N+C  
Line 273: unprotected comma in lemma  
ሥራገጣጣጣጣ .PART  
Line 274: unprotected comma in lemma  
ሥራገጣጣጣጣ .PRO+DEM  
Line 275: unprotected comma in lemma  
ሥራገጣጣጣጣ .PART  
Line 276: unprotected comma in lemma  
ሥራገጣጣጣጣ .PRO+DEM  
Line 277: unprotected comma in lemma  
ሥራገጣጣጣጣ .PART  
Line 278: unprotected comma in lemma  
ሥራገጣጣጣጣ .pe.  
Line 279: unprotected comma in lemma  
ሥራገጣጣጣጣ .PART  
Line 280: unprotected comma in lemma  
ሥራገጣጣጣጣ .PRO+PERS  
Line 282: unprotected comma in lemma  
ሥራገጣጣጣጣ .I+CONJ  
Line 283: unprotected comma in lemma  
ሥራገጣጣጣጣ .PRO+DEM  
Line 284: unprotected comma in lemma  
ሥራገጣጣጣጣ .PART  
Line 285: unprotected comma in lemma  
ሥራገጣጣጣጣ .A  
Line 289: unprotected comma in lemma  
ሥራገጣጣጣጣ .PART  
Line 290: unprotected comma in lemma  
ሥራገጣጣጣጣ .I+CONJ  
Line 293: unprotected comma in lemma  
ሥራገጣጣጣጣ .I+PREP  
Line 294: unprotected comma in lemma  
ሥራገጣጣጣጣ .PRO+PERS  
Line 299: unprotected comma in lemma  
ሥራገጣጣጣጣ .PRO+DEM  
Line 300: unprotected comma in lemma  
ሥራገጣጣጣጣ .I+PREP  
Line 301: unprotected comma in lemma  
ሥራገጣጣጣጣ .PART  
Line 302: unprotected comma in lemma  
ሥራገጣጣጣጣ .etpe.  
Line 303: unprotected comma in lemma  
ሥራገጣጣጣጣ .PART  
Line 304: unprotected comma in lemma  
ሥራገጣጣጣጣ .etpe.  
Line 311: unprotected comma in lemma  
ሥራገጣጣጣጣ .PART  
Line 312: unprotected comma in lemma  
ሥራገጣጣጣጣ .I+PREP  
Line 319: unprotected comma in lemma  
ሥራገጣጣጣጣ .I+PREP  
Line 320: unprotected comma in lemma  
ሥራገጣጣጣጣ .pe.  
Line 321: unprotected comma in lemma  
ሥራገጣጣጣጣ .I+PREP  
Line 322: unprotected comma in lemma  
ሥራገጣጣጣጣ .N+C  
Line 324: unprotected comma in lemma  
ሥራገጣጣጣጣ .pe.  
Line 325: unprotected comma in lemma  
ሥራገጣጣጣጣ .PRO+PERS  
Line 328: unprotected comma in lemma  
ሥራገጣጣጣጣ .I+CONJ  
Line 329: unprotected comma in lemma  
ሥራገጣጣጣጣ .I+ADV  
Line 330: unprotected comma in lemma  
ሥራገጣጣጣጣ .PART  
Line 331: unprotected comma in lemma  
ሥራገጣጣጣጣ .I+PREP  
Line 332: unprotected comma in lemma  
ሥራገጣጣጣጣ .N+P  
Line 333: unprotected comma in lemma  
ሥራገጣጣጣጣ .pe.  
Line 334: unprotected comma in lemma  
ሥራገጣጣጣጣ .PRO+PERS

Line 335: unprotected comma in lemma  
כחצו,כחצו,ד .I+PREP  
Line 336: unprotected comma in lemma  
כחצו,כחצו,ד .N+C  
Line 339: unprotected comma in lemma  
כחצו,כחצו,ד .etpe.  
Line 340: unprotected comma in lemma  
כחצו,כחצו,ד .PRO+PERS  
Line 341: unprotected comma in lemma  
כחצו,כחצו,ד .I+CONJ  
Line 342: unprotected comma in lemma  
כחצו,כחצו,ד .I+PREP  
Line 343: unprotected comma in lemma  
כחצו,כחצו,ד .PRO+PERS  
Line 344: unprotected comma in lemma  
כחצו,כחצו,ד .etpe.  
Line 345: unprotected comma in lemma  
כחצו,כחצו,ד .PRO+PERS  
Line 346: unprotected comma in lemma  
כחצו,כחצו,ד .etpe.  
Line 347: unprotected comma in lemma  
כחצו,כחצו,ד .PRO+PERS  
Line 348: unprotected comma in lemma  
כחצו,כחצו,ד .pe.  
Line 349: unprotected comma in lemma  
כחצו,כחצו,ד .PRO+PERS  
Line 350: unprotected comma in lemma  
כחצו,כחצו,ד .I+PREP  
Line 351: unprotected comma in lemma  
כחצו,כחצו,ד .PRO+PERS  
Line 352: unprotected comma in lemma  
כחצו,כחצו,ד .PART  
Line 353: unprotected comma in lemma  
כחצו,כחצו,ד .pe.  
Line 356: unprotected comma in lemma  
כחצו,כחצו,ד .I+CONJ  
Line 357: unprotected comma in lemma  
כחצו,כחצו,ד .I+ADV  
Line 358: unprotected comma in lemma  
כחצו,כחצו,ד .PART  
Line 359: unprotected comma in lemma  
כחצו,כחצו,ד .I+PREP  
Line 360: unprotected comma in lemma  
כחצו,כחצו,ד .PART  
Line 361: unprotected comma in lemma  
כחצו,כחצו,ד .A  
Line 365: unprotected comma in lemma  
כחצו,כחצו,ד .PART  
Line 366: unprotected comma in lemma  
כחצו,כחצו,ד .pe.  
Line 368: unprotected comma in lemma  
כחצו,כחצו,ד .PART  
Line 369: unprotected comma in lemma  
כחצו,כחצו,ד .pe.  
Line 374: unprotected comma in lemma  
כחצו,כחצו,ד .I+CONJ  
Line 375: unprotected comma in lemma  
כחצו,כחצו,ד .PRO+PERS  
Line 376: unprotected comma in lemma  
כחצו,כחצו,ד .PART  
Line 377: unprotected comma in lemma  
כחצו,כחצו,ד .N+C  
Line 381: unprotected comma in lemma  
כחצו,כחצו,ד .PART  
Line 382: unprotected comma in lemma  
כחצו,כחצו,ד .pe.  
Line 383: unprotected comma in lemma  
כחצו,כחצו,ד .I+PREP  
Line 384: unprotected comma in lemma  
כחצו,כחצו,ד .PRO+PERS  
Line 393: unprotected comma in lemma  
כחצו,כחצו,ד .PART  
Line 394: unprotected comma in lemma  
כחצו,כחצו,ד .PRO+PERS  
Line 395: unprotected comma in lemma  
כחצו,כחצו,ד .PART  
Line 396: unprotected comma in lemma  
כחצו,כחצו,ד .PRO+DEM  
Line 397: unprotected comma in lemma  
כחצו,כחצו,ד .PART  
Line 398: unprotected comma in lemma  
כחצו,כחצו,ד .N+C  
Line 399: unprotected comma in lemma  
כחצו,כחצו,ד .PRO+PERS

Line 400: unprotected comma in lemma  
 ጥዕና,ጥዕና,ጥዕና .PART  
 Line 401: unprotected comma in lemma  
 ጥዕና,ጥዕና,ጥዕና .N+C  
 Line 402: unprotected comma in lemma  
 ጥዕና,ጥዕና,ጥዕና .PART  
 Line 403: unprotected comma in lemma  
 ጥዕና,ጥዕና,ጥዕና .N+C  
 Line 404: unprotected comma in lemma  
 ጥዕና,ጥዕና,ጥዕና .PART  
 Line 405: unprotected comma in lemma  
 ጥዕና,ጥዕና,ጥዕና .N+C  
 Line 406: unprotected comma in lemma  
 ጥዕና,ጥዕና,ጥዕና .PART  
 Line 407: unprotected comma in lemma  
 ጥዕና,ጥዕና,ጥዕና .I+PREP  
 Line 408: unprotected comma in lemma  
 ጥዕና,ጥዕና,ጥዕና .PART

## Contenu du dictionnaire

ጥዕና,ጥዕና .PART  
 ጥዕና,ጥዕና .I+CONJ  
 ጥዕና,ጥዕና .I+PREP  
 ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና .I+ADV  
 ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና,ጥዕና .PART  
 ጥዕና,ጥዕና,ጥዕና .PART  
 ጥዕና,ጥዕና .N+P  
 ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና,ጥዕና .PART  
 ጥዕና,ጥዕና .pe.  
 ጥዕና,ጥዕና .pa.  
 ጥዕና,ጥዕና .I+PREP  
 ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና .pe.  
 ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና,ጥዕና .etpalp.  
 ጥዕና,ጥዕና .I+PREP  
 ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና .I+ADV  
 ጥዕና,ጥዕና .PRO+PERS  
 ጥዕና,ጥዕና,ጥዕና .I+CONJ  
 ጥዕና,ጥዕና,ጥዕና .I+ADV  
 ጥዕና,ጥዕና,ጥዕና .PART  
 ጥዕና,ጥዕና,ጥዕና .I+ADV  
 ጥዕና,ጥዕና .pe.  
 ጥዕና,ጥዕና .pe.  
 ጥዕና,ጥዕና .I+CONJ  
 ጥዕና,ጥዕና .pe.  
 ጥዕና,ጥዕና .I+PREP  
 ጥዕና,ጥዕና .PRO+DEM  
 ጥዕና,ጥዕና .A  
 ጥዕና,ጥዕና .I+CONJ  
 ጥዕና,ጥዕና .I+ADV  
 ጥዕና,ጥዕና .A  
 ጥዕና,ጥዕና .I+PREP  
 ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና .I+PREP  
 ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና,ጥዕና .I+CONJ  
 ጥዕና,ጥዕና,ጥዕና .I+PREP  
 ጥዕና,ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና .pe.  
 ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና,ጥዕና .I+PREP  
 ጥዕና,ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና .pe.  
 ጥዕና,ጥዕና,ጥዕና .N+C  
 ጥዕና,ጥዕና,ጥዕና .PRO+PERS  
 ጥዕና,ጥዕና .I+ADV  
 ጥዕና,ጥዕና,ጥዕና .PART  
 ጥዕና,ጥዕና .PART  
 ጥዕና,ጥዕና .pe.  
 ጥዕና,ጥዕና .PRO+INT  
 ጥዕና,ጥዕና .PRO+PERS





ከገጽገጽ .etpe.  
የገጽገጽ ቁጥር .NUM  
የገጽገጽ ቁጥር .PRO+PERS  
የገጽገጽ ቁጥር .NUM  
የገጽገጽ ቁጥር .N+C  
የገጽገጽ ቁጥር .pe.  
የገጽገጽ ቁጥር .I+PREP  
የገጽገጽ ቁጥር .NUM  
የገጽገጽ ቁጥር .N+C  
የገጽገጽ ቁጥር .etpe.  
የገጽገጽ ቁጥር .I+PREP  
የገጽገጽ ቁጥር .N+C  
የገጽገጽ ቁጥር .I+PREP  
የገጽገጽ ቁጥር .N+C  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .PRO+PERS  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .I+PREP  
የገጽገጽ ቁጥር .I+ADV  
የገጽገጽ ቁጥር .N+C  
የገጽገጽ ቁጥር .I+PREP  
የገጽገጽ ቁጥር .I+PREP  
የገጽገጽ ቁጥር .N+C  
የገጽገጽ ቁጥር .I+CONJ  
የገጽገጽ ቁጥር .N+C  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .I+PREP  
የገጽገጽ ቁጥር .I+ADV  
የገጽገጽ ቁጥር .N+C  
የገጽገጽ ቁጥር .I+CONJ  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .pe.  
የገጽገጽ ቁጥር .N+C  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .pa.  
የገጽገጽ ቁጥር .A  
የገጽገጽ ቁጥር .PRO+DEM  
የገጽገጽ ቁጥር .I+PREP  
የገጽገጽ ቁጥር .PRO+PERS  
የገጽገጽ ቁጥር .I+PREP  
የገጽገጽ ቁጥር .N+C  
የገጽገጽ ቁጥር .I+ADV  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .PRO+DEM  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .PRO+DEM  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .pe.  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .PRO+PERS  
የገጽገጽ ቁጥር .pe.  
የገጽገጽ ቁጥር .I+CONJ  
የገጽገጽ ቁጥር .PRO+DEM  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .A  
የገጽገጽ ቁጥር .pe.  
የገጽገጽ ቁጥር .pa.  
የገጽገጽ ቁጥር .I+ADV  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .I+CONJ  
የገጽገጽ ቁጥር .pe.  
የገጽገጽ ቁጥር .A  
የገጽገጽ ቁጥር .I+PREP  
የገጽገጽ ቁጥር .PRO+PERS  
የገጽገጽ ቁጥር .af.  
የገጽገጽ ቁጥር .af.  
የገጽገጽ ቁጥር .N+C  
የገጽገጽ ቁጥር .N+C  
የገጽገጽ ቁጥር .PRO+DEM  
የገጽገጽ ቁጥር .I+PREP  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .etpe.  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .etpe.  
የገጽገጽ ቁጥር .PRO+DEM  
የገጽገጽ ቁጥር .N+C  
የገጽገጽ ቁጥር .palp.  
የገጽገጽ ቁጥር .I+ADV  
የገጽገጽ ቁጥር .N+C  
የገጽገጽ ቁጥር .I+PREP  
የገጽገጽ ቁጥር .PART  
የገጽገጽ ቁጥር .I+PREP  
የገጽገጽ ቁጥር .pa.





ሥራ, ሥራ ላይ ስራ .N+C  
ሥራ, ሥራ ላይ ስራ .PRO+PERS  
ሥራ ላይ ስራ, ሥራ ላይ ስራ .PART  
ሥራ ላይ ስራ, ሥራ ላይ ስራ .N+C  
ሥራ ላይ ስራ, ሥራ ላይ ስራ .PART  
ሥራ ላይ ስራ, ሥራ ላይ ስራ .N+C  
ሥራ ላይ ስራ, ሥራ ላይ ስራ .PART  
ሥራ ላይ ስራ, ሥራ ላይ ስራ .I+PREP  
ሥራ ላይ ስራ, ሥራ ላይ ስራ .PART  
ሥራ ላይ ስራ, ሥራ ላይ ስራ .N+C