

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 référenciation;
- 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.

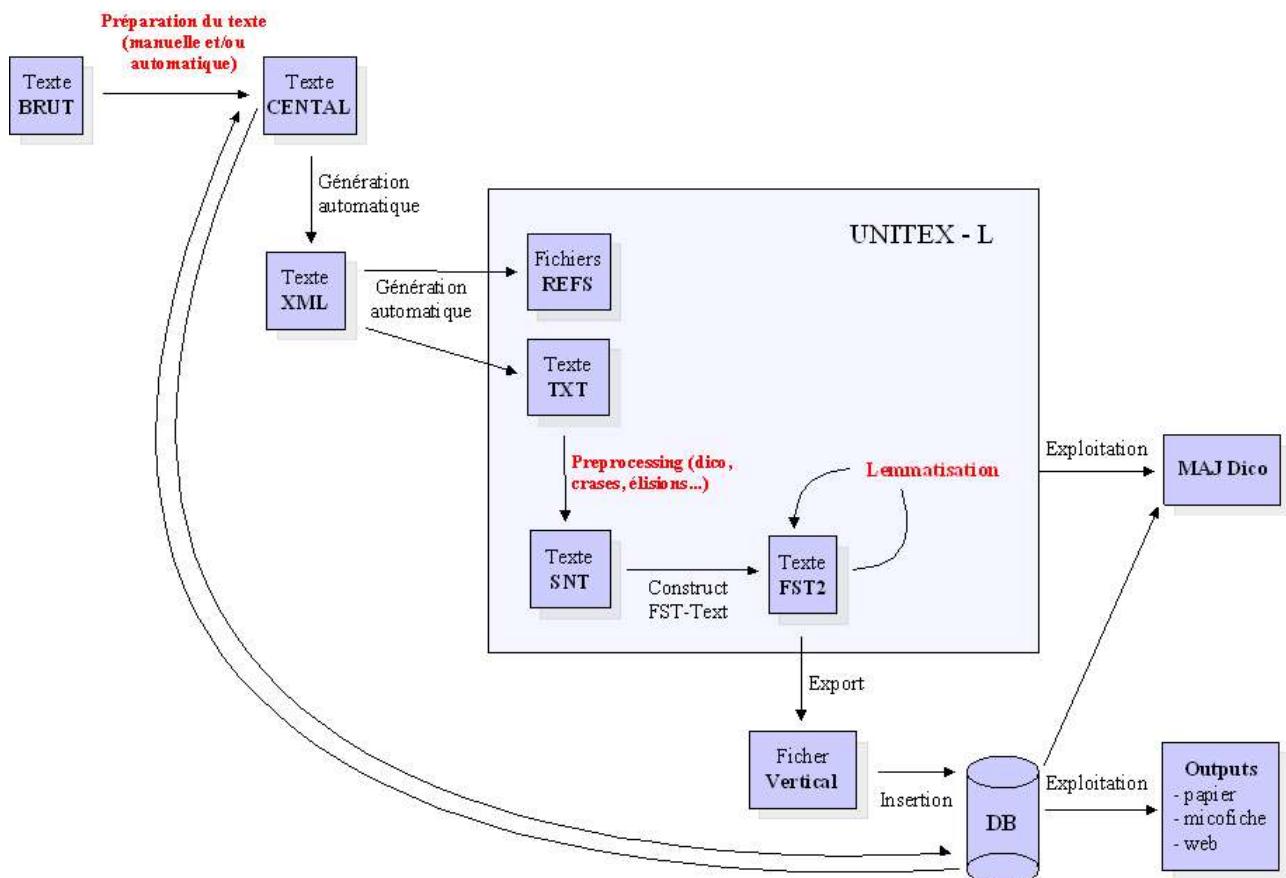
Le travail d'adaptation de ces logiciels au syriaque s'est fait avec l'aide de Sébastien Paumier.

L'adaptation du logiciel au syriaque a été envisagé 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 convertit 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 référenciation (\$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 convertit 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
KATA MAPKON
$0=Marc $1=1 $2=1 $9=1 $R=012
Αρχὴ τοῦ εὐαγγελίου Ἰησοῦ Χριστοῦ [μερικούς θεού].
$0=Marc $1=1 $2=2 $9=1 $R=012
Καθώς γέγραπται ἐν τῷ Ησαΐᾳ τῷ προφήτῃ,
[διού] ἀπόστελλω τὸν ἄγγελόν μου πρὸ προσώπου σου,
ὅς κατασκευάσει τὴν ὁδόν σου·
$0=Marc $1=1 $2=3 $9=1 $R=012
φωνῇ βοῶντος ἐν τῇ ἑρήμω.
Ἐτοιμάσατε τὴν ὁδὸν κυρίου,
εὐθείας ποιεῖτε τὰς τριβους αὐτοῦ—
$0=Marc $1=1 $2=4 $9=1 $R=012
ἐγένετο Ἰωάννης βαπτίζων ἐν τῇ ἑρήμω καὶ κηρύσσων
βάπτισμα μετανοίας εἰς ἄφεσιν ἀμαρτιῶν.
$0=Marc $1=1 $2=5 $9=1 $R=012
καὶ ἔξεπορεύετο
πρὸς αὐτὸν πᾶσα ἡ Ἰουδαϊα χώρα καὶ οἱ Ἱεροσολυμῖται
πάντες, καὶ ἐβαπτίζοντο ὑπὲρ αὐτοῦ ἐν τῷ Ἰορδάνῃ
πιπαμῷ ἐξομολογούμενοι τὰς ἀμαρτίας αὐτῶν.
$0=Marc $1=1 $2=6 $9=1 $R=012
καὶ ἦν
ὁ Ἰωάννης ἐνδεδημένος τρίχας καμῆλου καὶ ζώνην δερματίνην
περὶ τὴν ὄσφυν αὐτοῦ, καὶ ἔσθιων ἀκρίδας καὶ μέλι
ἄγριον.
$0=Marc $1=1 $2=7 $9=1 $R=012
καὶ ἐκήρυξσεν λέγων, Ἐρχεται ὁ ἴσχυρότερός
μου ὀπίσω μου, οὗ οὐκ εἰμὶ ίκανὸς κύψας λῦσαι τὸν
ιμάντα τῶν ὑποδημάτων αὐτοῦ.
```

```

<?xml version="1.0" encoding="UTF-16LE" ?>
<!DOCTYPE source="C:\TEMP\marcis.cental">
- <R0 utsShort="Evangelies" utsBreak="y" utsSelect="y">
- <R1 utsShort="Marc" utsBreak="y" utsSelect="y">
- <R2 utsShort="1" utsBreak="y" utsSelect="y">
- <R3 utsShort="t">
- <R9 utsShort="1">
  <![CDATA[ KATA MAPKON.
  ]]>
</R9>
- <R3 utsShort="1">
- <R9 utsShort="1">
  <![CDATA[ Αρχὴ τοῦ εὐαγγελίου Ἰησοῦ Χριστοῦ [μερικούς θεού].
  ]]>
</R9>
</R3>
- <R3 utsShort="2">
- <R9 utsShort="1">
  <![CDATA[ Καθὼς γέγραπται ἐν τῷ Ησαΐᾳ προφήτῃ
  ]]>
</R9>
- <R9 utsShort="2">
  <![CDATA[ Διού ἀπόστελλω τὸν ἄγγελόν μου πρὸ προσώπου σου,
  ]]>
</R9>
- <R9 utsShort="3">
  <![CDATA[ Ος κατασκευάσει τὸν οδὸν σου·
  ]]>
</R9>
</R3>
- <R3 utsShort="3">
- <R9 utsShort="1">
  <![CDATA[ φωνῇ βοῶντος ον τῷ ορθῷ,
  ]]>
</R9>
- <R9 utsShort="2">
  <![CDATA[ Ετοιμάσατε τὸν οδὸν κυρίου,
  ]]>
</R9>
- <R9 utsShort="3">
  <![CDATA[ εὐθείας ποιεῖτε τῷς τριβους αὐτοῦ —
  ]]>
</R9>
</R3>
- <R3 utsShort="4">
- <R9 utsShort="1">
  <![CDATA[ Οὐθετο οὐδενῆς βεπτούσων ον τῷ ορθῷ καθηρίσσων
  ]]>
</R9>
```

Texte XML

Texte au format CENTAL

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 adapté 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 Unicode¹ 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 SIL² pour leur système Graphite³, contact avec George Kiraz de l'organisation de Beth Mardutho⁴) 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

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

version 1.1. Les modifications, apportées avec le concours de Sébastien Paumier pour faire fonctionner la recherche et améliorer l'affichage des scripts « de droite à gauche », ont été effectuées sur la dernière version d'Unitex (1.2). Nous avons donc travaillé simultanément sur deux états différents de l'interface. Cette situation peut être résolue par deux moyens. Le travail qui a été effectué sur la version 1.2 pour l'amélioration du support du syriaque pourrait être également mené sur la version 1.1 (mais cela dépassera probablement le simple copier/coller, étant donné les changements intervenus entre la version 1.1 et la version 1.2). La seconde possibilité est de porter l'interface de lemmatisation vers la version 1.2 d'Unitex, ce qui implique un travail non négligeable à divers niveaux. Aucune décision n'a été prise sur l'option à privilégier.

Globalement...

A ce stade, l'intégration du syriaque à l'interface de lemmatisation (UnitexL) telle qu'elle avait été construite pour le grec ancien n'est pas effective. Par contre, le support sous Unitex des langages « de droite à gauche » en général, et du syriaque et de l'arabe en particulier, aura été grandement amélioré. De plus, des versions expérimentales des diverses ressources linguistiques ont déjà été construites et sont opérationnelles.

2. Ressources

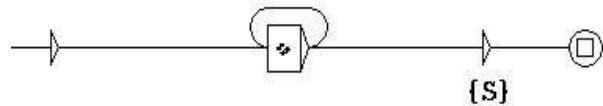
Alphabet

Le fichier alphabet définit les caractères de la langue.

二十一
卷之三

Graphe de découpe en phrases

Ce graphe permet un découpage du texte en phrases. Il s'agit bien entendu d'une version très simple, mais qui a déjà permis de fractionner le texte dont nous disposons en 11 phrases. Ce graphe doit bien entendu être complexifié afin de prendre en compte tous les cas se retrouvant dans les textes syriaques.



Catégories grammaticales

Cette liste a été construite par A.Schmidt et J.-C. Haelewyck (ref. PAYNE-SMITH).

N	Noms
N+C	Noms communs
N+P	Noms propres
N+P+ANT	Noms propres anthroponymiques
N+P+PATR	Noms propres patronymiques
N+P+ETH	Noms propres ethniques
N+P+TOP	Noms propres toponymiques
A	Adjectifs
V+chiffres	Verbes
PRO	Pronoms
PRO+PERS	Pronoms personnels
PRO+DEM	Pronoms démonstratifs
PRO+IND	Pronoms indéfinis
PRO+INT	Pronoms interrogatifs
PRO+REL	Pronoms relatifs
NUM	Numéraux
NUM+CAR	Numéraux cardinaux
NUM+ORD	Numéraux ordinaux
I	Invariables
I+ADV	Invariables adverbes
I+CONJ	Invariables conjonction
I+NEG	Invariables négation
I+INTERJ	Invariables interjection
I+PREP	Invariables préposition

Le DELA Syriaque

Rapport de compatibilité

Line 8: unprotected comma in lemma
 ↗ .PART

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

Line 190: unprotected comma in lemma
نَحْنُ، أَنْتَ، إِنْتَ .I+CONJ
Line 191: unprotected comma in lemma
نَحْنُ، أَنْتَ، إِنْتَ .I+ADV
Line 193: unprotected comma in lemma
مُلْحَمٌ، لَهُ .I+CONJ
Line 194: unprotected comma in lemma
مُلْحَمٌ، لَهُ .I+PREP
Line 195: unprotected comma in lemma
مُلْحَمٌ، لَهُ .N+C
Line 205: unprotected comma in lemma
نَمَّ، نَمَّ .PART
Line 206: unprotected comma in lemma
نَمَّ، نَمَّ .pe.
Line 207: unprotected comma in lemma
نَمَّ، نَمَّ .PART
Line 208: unprotected comma in lemma
نَمَّ، نَمَّ .I+PREP
Line 209: unprotected comma in lemma
نَمَّ، نَمَّ .N+C
Line 213: unprotected comma in lemma
نَمَّ، نَمَّ .PART
Line 214: unprotected comma in lemma
نَمَّ، نَمَّ .I+PREP
Line 216: unprotected comma in lemma
نَمَّ، نَمَّ .I+CONJ
Line 217: unprotected comma in lemma
نَمَّ، نَمَّ .I+PREP
Line 222: unprotected comma in lemma
نَمَّ، نَمَّ .I+PREP
Line 223: unprotected comma in lemma
نَمَّ، نَمَّ .N+C
Line 225: unprotected comma in lemma
نَمَّ، نَمَّ .PART
Line 226: unprotected comma in lemma
نَمَّ، نَمَّ .I+ADV
Line 231: unprotected comma in lemma
نَمَّ، نَمَّ .NUM
Line 232: unprotected comma in lemma
نَمَّ، نَمَّ .PRO+PERS
Line 236: unprotected comma in lemma
نَمَّ، نَمَّ .I+PREP
Line 237: unprotected comma in lemma
نَمَّ، نَمَّ .NUM
Line 240: unprotected comma in lemma
نَمَّ، نَمَّ .I+PREP
Line 241: unprotected comma in lemma
نَمَّ، نَمَّ .N+C
Line 242: unprotected comma in lemma
نَمَّ، نَمَّ .I+PREP
Line 243: unprotected comma in lemma
نَمَّ، نَمَّ .N+C
Line 244: unprotected comma in lemma
نَمَّ، نَمَّ .PART
Line 245: unprotected comma in lemma
نَمَّ، نَمَّ .PRO+PERS
Line 246: unprotected comma in lemma
نَمَّ، نَمَّ .PART
Line 247: unprotected comma in lemma
نَمَّ، نَمَّ .I+PREP
Line 248: unprotected comma in lemma
نَمَّ، نَمَّ .I+ADV
Line 249: unprotected comma in lemma
نَمَّ، نَمَّ .N+C
Line 250: unprotected comma in lemma
نَمَّ، نَمَّ .I+PREP
Line 253: unprotected comma in lemma
نَمَّ، نَمَّ .I+CONJ
Line 254: unprotected comma in lemma
نَمَّ، نَمَّ .N+C
Line 255: unprotected comma in lemma
نَمَّ، نَمَّ .PART
Line 256: unprotected comma in lemma
نَمَّ، نَمَّ .I+PREP
Line 257: unprotected comma in lemma
نَمَّ، نَمَّ .I+ADV
Line 258: unprotected comma in lemma
نَمَّ، نَمَّ .N+C
Line 259: unprotected comma in lemma
نَمَّ، نَمَّ .I+CONJ
Line 260: unprotected comma in lemma
نَمَّ، نَمَّ .PART
Line 264: unprotected comma in lemma
نَمَّ، نَمَّ .PART

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

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

Contenu du dictionnaire

تَرَكَ, تَرَكَ .PART
 تَرَكَ+CONJ
 حَلَّا .I+PREP
 حَلَّا صَفَّا N+C
 حَلَّا صَفَّا N+C
 حَدَّا أَنْ, حَدَّا أَنْ .I+ADV
 حَدَّا, حَدَّا N+C
 حَدَّا, حَدَّا .PART
 حَدَّا, حَدَّا, حَدَّا .PART
 حَدَّا, حَدَّا .N+P
 حَدَّا, حَدَّا N+C
 حَدَّا, حَدَّا, حَدَّا N+C
 حَدَّا, حَدَّا, حَدَّا .PART
 حَدَّا, حَدَّا etpe.
 حَدَّا, حَدَّا .pa.
 حَدَّا, حَدَّا .I+PREP
 حَدَّا, حَدَّا .N+C
 حَدَّا, حَدَّا .pe.
 حَدَّا, حَدَّا .N+C
 حَدَّا, حَدَّا etpalp.
 حَدَّا, حَدَّا .I+PREP
 حَدَّا, حَدَّا .N+C
 حَدَّا, حَدَّا .I+ADV
 حَدَّا, حَدَّا .PRO+PERS
 حَدَّا, حَدَّا .I+CONJ
 حَدَّا, حَدَّا .I+ADV
 حَدَّا, حَدَّا .PART
 حَدَّا, حَدَّا .I+ADV
 حَدَّا, حَدَّا .pe.
 حَدَّا, حَدَّا .pe.
 حَدَّا, حَدَّا .I+CONJ
 حَدَّا, حَدَّا .pe.
 حَدَّا, حَدَّا .I+PREP
 حَدَّا, حَدَّا .PRO+DEM
 حَدَّا, حَدَّا .A
 حَدَّا, حَدَّا .I+CONJ
 حَدَّا, حَدَّا .I+ADV
 حَدَّا, حَدَّا .A
 حَدَّا, حَدَّا .I+PREP
 حَدَّا, حَدَّا .N+C
 حَدَّا, حَدَّا .I+PREP
 حَدَّا, حَدَّا .N+C
 حَدَّا, حَدَّا, حَدَّا .I+CONJ
 حَدَّا, حَدَّا, حَدَّا .I+PREP
 حَدَّا, حَدَّا, حَدَّا .N+C
 حَدَّا, حَدَّا .pe.
 حَدَّا, حَدَّا .N+C
 حَدَّا, حَدَّا, حَدَّا .I+PREP
 حَدَّا, حَدَّا .N+C
 حَدَّا, حَدَّا .N+C
 حَدَّا, حَدَّا .N+C
 حَدَّا, حَدَّا .N+C
 حَدَّا, حَدَّا .pe.
 حَدَّا, حَدَّا etpalp .N+C
 حَدَّا, حَدَّا, حَدَّا .PRO+PERS
 حَدَّا, حَدَّا .I+ADV
 حَدَّا, حَدَّا .PART
 حَدَّا, حَدَّا .PART
 حَدَّا, حَدَّا .pe.
 حَدَّا, حَدَّا .PRO+INT
 حَدَّا, حَدَّا .PRO+PERS

حَتَّىٰ .I+CONJ
 لِكُلِّ .N+C
 أَنْ, أَنْهَا .etpe.
 إِنْ, إِنْهَا .I+CONJ
 إِنْ, إِنْهَا .etpe.
 كُلُّ, كُلُّهُ .N+C
 كُلُّ, كُلُّهُ .PRO+PERS
 كُلُّ, كُلُّهُ .N+C
 كُلُّ, كُلُّهُ .PART
 كُلُّ, كُلُّهُ .N+C
 كُلُّ, كُلُّهُ .A
 كُلُّ, كُلُّهُ .PRO+DEM
 كُلُّ, كُلُّهُ .PART
 كُلُّ, كُلُّهُ .N+C
 كُلُّ, كُلُّهُ .N+P
 كُلُّ, كُلُّهُ .af.
 كُلُّ, كُلُّهُ .pa.
 كُلُّ, كُلُّهُ .N+C
 كُلُّ, كُلُّهُ .PRO+PER
 كُلُّ, كُلُّهُ .PART
 كُلُّ, كُلُّهُ .N+C
 كُلُّ, كُلُّهُ .PRO+PERS
 كُلُّ, كُلُّهُ .I+CONJ
 كُلُّ, كُلُّهُ .af.
 كُلُّ, كُلُّهُ .N+C
 كُلُّ, كُلُّهُ .PRO+PERS
 كُلُّ, كُلُّهُ .PART
 كُلُّ, كُلُّهُ .N+C
 كُلُّ, كُلُّهُ .etpa.
 كُلُّ, كُلُّهُ .N+C
 كُلُّ, كُلُّهُ .etpa.
 كُلُّ, كُلُّهُ .N+C
 كُلُّ, كُلُّهُ .A
 كُلُّ, كُلُّهُ .etpe.
 كُلُّ, كُلُّهُ .N+C
 كُلُّ, كُلُّهُ .pa.
 كُلُّ, كُلُّهُ .N+C
 كُلُّ, كُلُّهُ .PRO+PERS
 كُلُّ, كُلُّهُ .PART
 كُلُّ, كُلُّهُ .N+C
 كُلُّ, كُلُّهُ .PART
 كُلُّ, كُلُّهُ .N+C
 كُلُّ, كُلُّهُ .I+ADV
 كُلُّ, كُلُّهُ .I+CONJ
 كُلُّ, كُلُّهُ .I+ADV
 كُلُّ, كُلُّهُ .PRO+PERS
 مُكْلِمٌ, مُكْلِمٌ .I+CONJ
 مُكْلِمٌ, مُكْلِمٌ .I+PREP
 مُكْلِمٌ, مُكْلِمٌ .I+PREP
 مُكْلِمٌ, مُكْلِمٌ .N+C
 مُكْلِمٌ, مُكْلِمٌ .A
 مُكْلِمٌ, مُكْلِمٌ .etpe.
 مُكْلِمٌ, مُكْلِمٌ .A
 مُكْلِمٌ, مُكْلِمٌ .pa.
 مُكْلِمٌ, مُكْلِمٌ .A
 مُكْلِمٌ, مُكْلِمٌ .pa.
 مُكْلِمٌ, مُكْلِمٌ .I+ADV
 مُكْلِمٌ, مُكْلِمٌ .paiel
 مُكْلِمٌ, مُكْلِمٌ .I+ADV
 مُكْلِمٌ, مُكْلِمٌ .PART
 مُكْلِمٌ, مُكْلِمٌ .pe.
 مُكْلِمٌ, مُكْلِمٌ .PART
 مُكْلِمٌ, مُكْلِمٌ .I+PREP
 مُكْلِمٌ, مُكْلِمٌ .N+C
 مُكْلِمٌ, مُكْلِمٌ .pe.
 مُكْلِمٌ, مُكْلِمٌ .I+CONJ
 مُكْلِمٌ, مُكْلِمٌ .I+CONJ
 مُكْلِمٌ, مُكْلِمٌ .PART
 مُكْلِمٌ, مُكْلِمٌ .I+PREP
 مُكْلِمٌ, مُكْلِمٌ .N+C(qattolo)
 مُكْلِمٌ, مُكْلِمٌ .pe.
 مُكْلِمٌ, مُكْلِمٌ .PRO+DEM
 مُكْلِمٌ, مُكْلِمٌ .I+CONJ
 مُكْلِمٌ, مُكْلِمٌ .I+PREP
 مُكْلِمٌ, مُكْلِمٌ .N+C
 مُكْلِمٌ, مُكْلِمٌ .I+CONJ
 مُكْلِمٌ, مُكْلِمٌ .PART
 مُكْلِمٌ, مُكْلِمٌ .I+ADV
 مُكْلِمٌ, مُكْلِمٌ .N+C
 مُكْلِمٌ, مُكْلِمٌ .N+C

تَوْرِيزْتَهُ .etpe.
تَوْرِيزْتَهُمْ .NUM
تَوْرِيزْتَهُمْ .PRO+PERS
تَوْرِيزْتَهُ .NUM
تَوْرِيزْتَهُ .N+C
تَوْرِيزْتَهُ .pe.
تَوْرِيزْتَهُ .I+PREP
تَوْرِيزْتَهُ .NUM
تَوْرِيزْتَهُ .N+C
تَوْرِيزْتَهُ .etpe.
تَوْرِيزْتَهُ .I+PREP
تَوْرِيزْتَهُ .N+C
تَوْرِيزْتَهُ .I+PREP
تَوْرِيزْتَهُ .N+C
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .PRO+PERS
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .I+PREP
تَوْرِيزْتَهُ .I+ADV
تَوْرِيزْتَهُ .N+C
تَوْرِيزْتَهُ .I+PREP
تَوْرِيزْتَهُ .N+C
تَوْرِيزْتَهُ .I+CONJ
تَوْرِيزْتَهُ .N+C
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .I+PREP
تَوْرِيزْتَهُ .I+ADV
تَوْرِيزْتَهُ .N+C
تَوْرِيزْتَهُ .I+CONJ
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .pe.
تَوْرِيزْتَهُ .N+C
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .pa.
تَوْرِيزْتَهُ .A
تَوْرِيزْتَهُ .PRO+DEM
تَوْرِيزْتَهُ .I+PREP
تَوْرِيزْتَهُ .PRO+PERS
تَوْرِيزْتَهُ .I+PREP
تَوْرِيزْتَهُ .N+C
تَوْرِيزْتَهُ .I+ADV
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .PRO+DEM
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .PRO+DEM
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .I+ADV
تَوْرِيزْتَهُ .pe.
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .I+PREP
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .I+CONJ
تَوْرِيزْتَهُ .PRO+DEM
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .A
تَوْرِيزْتَهُ .pe.
تَوْرِيزْتَهُ .pa.
تَوْرِيزْتَهُ .I+ADV
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .I+CONJ
تَوْرِيزْتَهُ .pe.
تَوْرِيزْتَهُ .A
تَوْرِيزْتَهُ .I+PREP
تَوْرِيزْتَهُ .PRO+PERS
تَوْرِيزْتَهُ .af.
تَوْرِيزْتَهُ .af.
تَوْرِيزْتَهُ .af.
تَوْرِيزْتَهُ .N+C
تَوْرِيزْتَهُ .N+C
تَوْرِيزْتَهُ .PRO+DEM
تَوْرِيزْتَهُ .I+PREP
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .etpe.
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .etpe.
تَوْرِيزْتَهُ .PRO+DEM
تَوْرِيزْتَهُ .N+C
تَوْرِيزْتَهُ .palp.
تَوْرِيزْتَهُ .I+ADV
تَوْرِيزْتَهُ .N+C
تَوْرِيزْتَهُ .I+PREP
تَوْرِيزْتَهُ .PART
تَوْرِيزْتَهُ .I+PREP
تَوْرِيزْتَهُ .pa.

اَنْ .+CONJ
 اَنْ, اَنْ .pe.
 اَنْ, اَنْ .pe.
 اَنْ, اَنْ .+ADV
 اَنْ, اَنْ .A
 اَنْ, اَنْ .I+PREP
 اَنْ, اَنْ .pe.
 اَنْ, اَنْ .I+PREP
 اَنْ, اَنْ .N+C
 اَنْ, اَنْ .A
 اَنْ, اَنْ .A
 اَنْ, اَنْ .pe.
 اَنْ, اَنْ .PRO+PERS
 اَنْ, اَنْ .A
 اَنْ, اَنْ .pe.
 اَنْ, اَنْ .I+CONJ
 اَنْ, اَنْ .+ADV
 اَنْ, اَنْ .PART
 اَنْ, اَنْ .I+PREP
 اَنْ, اَنْ .N+P
 اَنْ, اَنْ .pe.
 اَنْ, اَنْ .PRO+PERS
 اَنْ, اَنْ .I+PREP
 اَنْ, اَنْ .N+C
 اَنْ, اَنْ .pe.
 اَنْ, اَنْ .I+PREP
 اَنْ, اَنْ .etpe.
 اَنْ, اَنْ .PRO+PERS
 اَنْ, اَنْ .I+CONJ
 اَنْ, اَنْ .I+PREP
 اَنْ, اَنْ .PRO+PERS
 اَنْ, اَنْ .etpe.
 اَنْ, اَنْ .PRO+PERS
 اَنْ, اَنْ .etpe.
 اَنْ, اَنْ .PRO+PERS
 اَنْ, اَنْ .etpe.
 اَنْ, اَنْ .PRO+PERS
 اَنْ, اَنْ .pe.
 اَنْ, اَنْ .PRO+PERS
 اَنْ, اَنْ .I+PREP
 اَنْ, اَنْ .I+PREP
 اَنْ, اَنْ .PRO+PERS
 اَنْ, اَنْ .PART
 اَنْ, اَنْ .pe.
 اَنْ, اَنْ .N+C(qetlo)
 اَنْ, اَنْ .A
 اَنْ, اَنْ .I+CONJ
 اَنْ, اَنْ .+ADV
 اَنْ, اَنْ .PART
 اَنْ, اَنْ .I+PREP
 اَنْ, اَنْ .PART
 اَنْ, اَنْ .A
 اَنْ, اَنْ .A
 اَنْ, اَنْ .pe.
 اَنْ, اَنْ .A
 اَنْ, اَنْ .PART
 اَنْ, اَنْ .pe.
 اَنْ, اَنْ .pe.
 اَنْ, اَنْ .PART
 اَنْ, اَنْ .pe.
 اَنْ, اَنْ .I+ADV
 اَنْ, اَنْ .I+ADV
 اَنْ, اَنْ .PART
 اَنْ, اَنْ .A
 اَنْ, اَنْ .N+C
 اَنْ, اَنْ .I+CONJ
 اَنْ, اَنْ .PRO+PERS
 اَنْ, اَنْ .PART
 اَنْ, اَنْ .N+C
 اَنْ, اَنْ .pa.
 اَنْ, اَنْ .I+ADV
 اَنْ, اَنْ .I+ADV
 اَنْ, اَنْ .I+ADV
 اَنْ, اَنْ .I+ADV
 اَنْ, اَنْ .N+C
 اَنْ, اَنْ .PART
 اَنْ, اَنْ .PRO+PERS
 اَنْ, اَنْ .PART
 اَنْ, اَنْ .PRO+DEM
 اَنْ, اَنْ .PART

نَاهِيَةٌ .N+C
نَاهِيَةٌ .PRO+PERS
نَاهِيَةٌ .PART
نَاهِيَةٌ .N+C
نَاهِيَةٌ .PART
نَاهِيَةٌ .N+C
نَاهِيَةٌ .PART
نَاهِيَةٌ .N+C
نَاهِيَةٌ .PART
نَاهِيَةٌ .I+PREP
نَاهِيَةٌ .PART
نَاهِيَةٌ .N+C