

Construire des grammaires de levée d'ambiguïtés pour Intex

Anne DISTER
Université de Liège

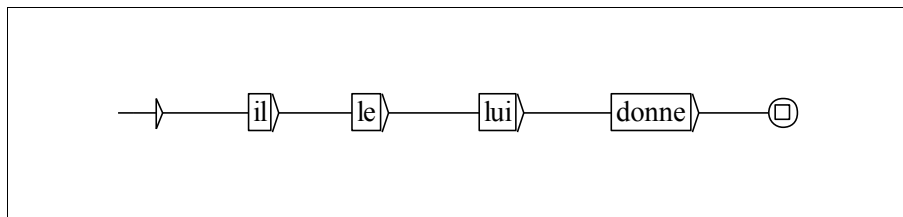
Introduction

Dans un premier temps, nous verrons les « règles » à connaître pour construire une grammaire locale de levée d'ambiguïtés; dans une seconde partie, nous détaillerons la manière dont les grammaires d'Intex s'appliquent sur un texte, et le type de résultats que l'on peut obtenir.

1. Construire des grammaires de levée d'ambiguïtés

1.1. Un automate

Les grammaires locales habituellement construites par les utilisateurs d'Intex, qui sont des grammaires de reconnaissance, ont la forme suivante :

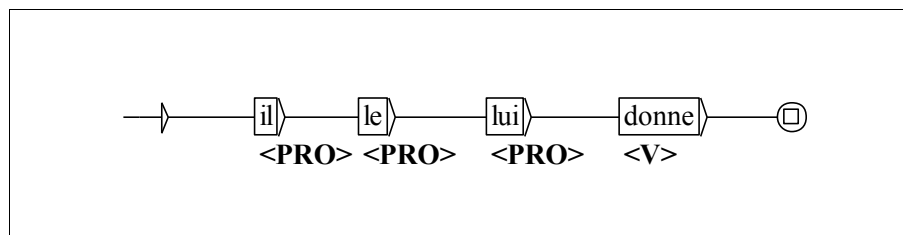


Grappe1

Il s'agit en fait d'un automate. Une séquence est reconnue par un automate si, partant du nœud initial, nous arrivons au nœud terminal en passant par tous les nœuds de l'automate (**on lit de gauche à droite**). Ainsi, Grappe1 reconnaît dans un texte la séquence *il le lui donne*. Mais des séquences telles que *il le donne*, *il lui donne* ou *elle le lui donne* ne seront pas reconnues.

1.2. Un transducteur

Les grammaires utilisées dans le cadre de la levée d'ambiguïtés sont quelque peu différentes : il s'agit de transducteurs.



Grappe2

Un transducteur est un automate à deux bandes : une bande de lecture qui reconnaît une séquence dans un texte, et une bande d'écriture qui associe des informations à cette séquence préalablement reconnue. Le graphe2, ayant reconnu dans un texte la séquence *il le lui donne*, associe à chaque mot de cette séquence l'information présente sous la boîte¹, c'est-à-dire la contrainte.

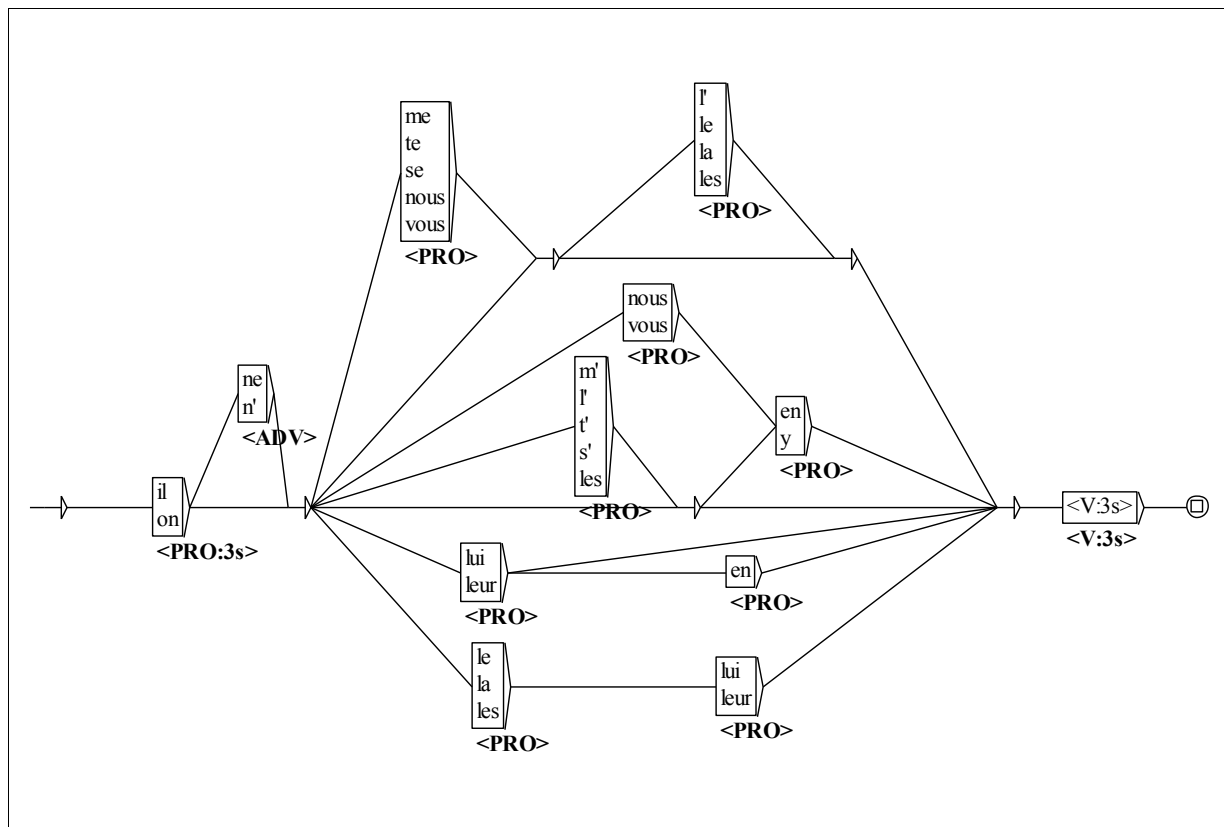
Dans un étiquetage traditionnel (nous reviendrons plus loin sur cette notion), le texte prend alors la forme suivante :

{il,PRO:3ms}{le,PRO:3ms}{lui,PRO:3s}{donne,donner.V:P1s:P3s:S1s:S3s:Y2s}.

Dans les boîtes, on peut trouver :

- la forme telle qu'elle apparaît dans le texte
- un lemme, noté entre crochets : <lemme>
- une catégorie, notée entre crochets et ne majuscules : <CAT>
- le symbole vide : <E>

Voici le graphe plus complet, dont Graphe2 n'est qu'une partie, qui traite toute la séquence entre les pronoms personnels sujet *il* ou *on* et le verbe.



Graphe3

1.3. Les contraintes

Dans la séquence qui précède, *il* est un mot non ambigu et la contrainte <PRO> peut donc sembler superflue. Or, il faut savoir que pour qu'une grammaire de levée d'ambiguïtés

¹ Boîte et nœud sont synonymes

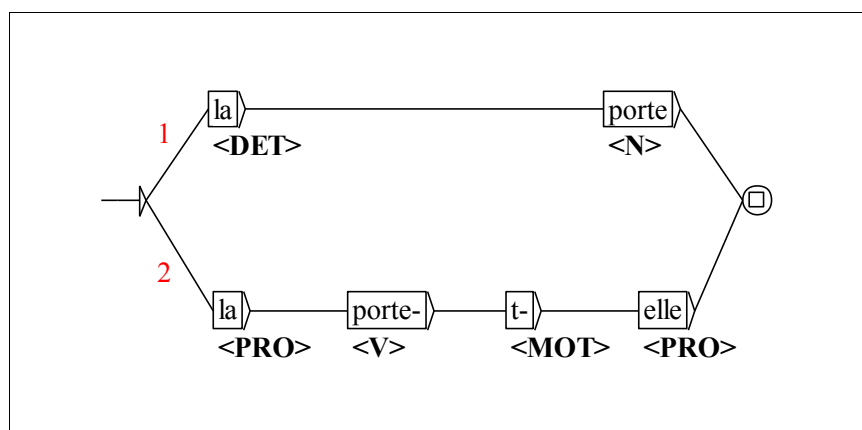
fonctionne, **une contrainte est indispensable sous chaque nœud**². Cette contrainte peut être très générale : <MOT> suffit. Elle impose au mot reconnu d'être un mot : si ce mot est ambigu, il n'y a pas de levée d'ambiguïtés. Si le mot est non ambigu, il sera étiqueté dans le texte.

Il faut aussi noter, comme c'est le cas dans Graphe2, **qu'une contrainte peut être moins complète que l'information morphologique associée à l'entrée lexicale (informations du dictionnaire) de la forme que l'on veut désambiguïser**. Ainsi, la contrainte <PRO> suffit à désambiguïser *le* dont l'entrée lexicale dans le delafm est <le,PRO:3ms>. On aurait également pu écrire <PRO:3s> ou <PRO:3m> ou encore <PRO:ms>. **L'important est que les informations de la contrainte ne soient pas plus complètes que les informations de l'entrée lexicale** : <PRO:3ms+Hum> créerait une incohérence, et la séquence ne serait pas reconnue.

Ce qui vient d'être dit ici pour les contraintes est également valable pour les étiquettes dans les boîtes.

1.4. Plusieurs chemins peuvent être actifs en même temps

Lors de la levée d'ambiguïtés, plusieurs chemins peuvent être actifs en même temps, soit à l'intérieur de la même grammaire, soit dans deux grammaires appliquées simultanément. Lorsque deux chemins sont en concurrence, **c'est le chemin le plus long (celui qui contient le plus de nœuds) qui est reconnu**.

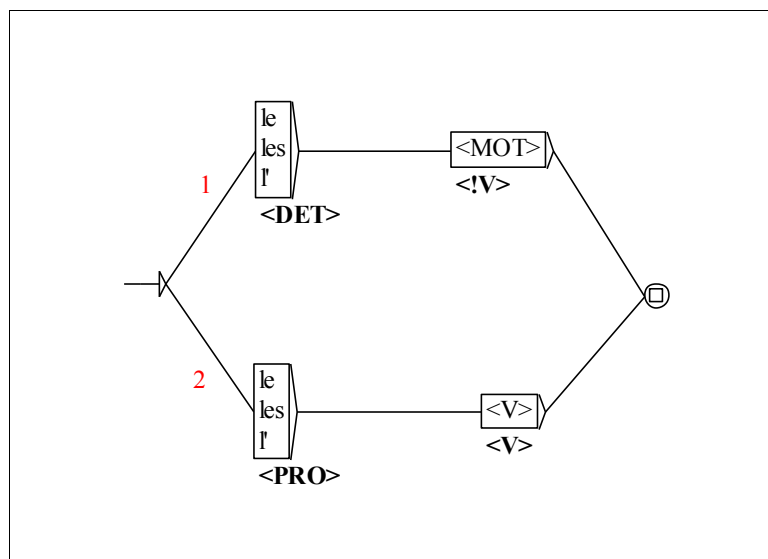


Graphe4

Si le texte est (...) *vers la porte de ma chambre* (...), c'est le chemin 1 de Graphe4 qui sera suivi, puisque lui seul permet, partant du nœud initial, de passer par tous les nœuds pour arriver au nœud terminal. Par contre, si le texte est *Maïa la porte-t-elle dans cœur ?*, c'est le chemin 2 qui sera suivi puisqu'il permet, toujours en partant du nœud initial, en arrivant au nœud terminal et en passant par tous les nœuds, de reconnaître une séquence plus longue. Dans le premier cas, la séquence *la porte* est analysée *DET+N*, dans le second cas, *PRO+V*.

Mais il peut arriver que les chemins soient de longueur égale, comme dans Graphe5.

² Si chaque nœud ne possède pas sa contrainte, la séquence n'est pas reconnue, donc pas désambiguïcée. Dans l'exemple présenté « il le lui donne », seul *il* sera étiqueté puisque non ambigu.



Graphe5

Le chemin 1 dit que les formes *le*, *les* et *l'* sont des déterminants et le mot qui suit n'est pas un verbe ; le chemin 2 impose aux formes d'être des pronoms et au verbe qui suit d'être un verbe.

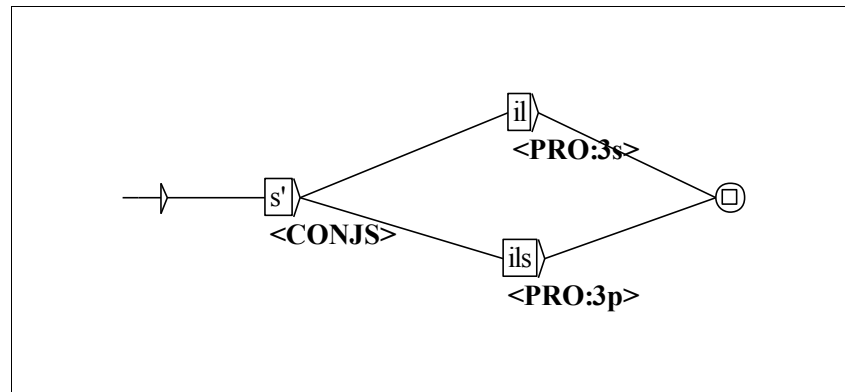
Si le mot qui suit *<le+les+l'>* ne peut être qu'un verbe, c'est le chemin 2 qui sera appliqué : c'est le cas par exemple de séquences comme *le baptisera*, *l'aime*, *les organisent* où la seule hypothèse pour *baptisera*, *aime* et *organisent* dans le *delafm* est V. Si ce mot ne peut jamais être un verbe (cette hypothèse ne se trouve pas dans l'entrée lexicale de la forme), c'est alors le chemin 1 qui est suivi : c'est le cas de *l'amiral*, *les belles*, *le très intelligent*, etc. Mais si le mot qui suit *<le+les+l'>* peut à la fois être un verbe et ne pas être un verbe, les 2 chemins sont actifs en même temps. C'est le cas de séquences comme *le combat*, *les élèves*, *l'écoute*, *le dîner*, etc. Or, **quand deux chemins de longueur égale sont actifs en même temps**³, **aucun des deux n'est préféré à l'autre**⁴. Dans ce cas, lors de l'étiquetage du texte, rien ne se produit. Nous reviendrons sur cette grammaire dans la seconde partie lorsque nous verrons que la levée d'ambiguïtés ne concerne pas uniquement l'étiquetage de texte.

1.5. Les grammaires s'appliquent de gauche à droite sur un texte

La grammaire suivante lève l'ambiguïté du *s'* lorsqu'il est suivi des pronoms *il* ou *ils*.

³ Nous parlons ici de chemins qui commencent au même endroit dans le texte.

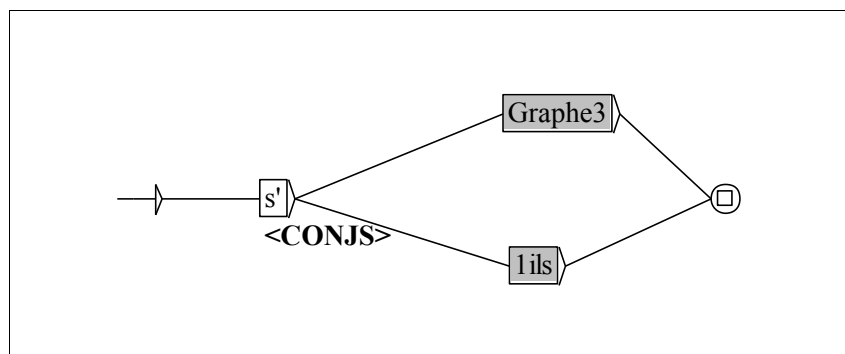
⁴ Ici, les deux chemins ont la même « allure » : une forme dans le premier nœud et une catégorie grammaticale dans le second. On pourrait imaginer que lorsque les chemins sont de même longueur, la priorité serait donnée à celui dont les contraintes sont les plus complètes ou dont les nœuds contiennent des mots plutôt que des catégories, etc. Or, aucune considération de ce type n'intervient.



Graphe6

Nous avons dit qu'il était fréquent que plusieurs grammaires s'appliquent simultanément sur un texte. Imaginons que ce soit le cas de Graphe6 et de Graphe3 sur la séquence « s'il le lui donne ». Les grammaires s'appliquent de gauche à droite sur le texte : c'est donc d'abord Graphe6 qui va reconnaître « s'il ». Cette séquence, reconnue, est alors écartée (puisque déjà analysée), et il reste la séquence « le lui donne ». Or, celle-ci ne peut être reconnue par Graphe3. En effet, avant d'arriver à *le*, Graphe3 doit rencontrer sur son chemin *il* ou *on*, ce qui n'est pas le cas ici.

Pour pallier ce type d'inconvénient, Intex offre la possibilité de construire des **graphes enchâssés** : une grammaire appelle dans son chemin une autre grammaire (le nœud est alors grisé⁵). Graphe7 remplace efficacement l'union de Graphe3 et de Graphe6 :



Graphe7

Après avoir rencontré *s'*, Graphe7 poursuit son chemin et reconnaît la séquence *il le lui donne* grâce à Graphe3 imbriqué⁶.

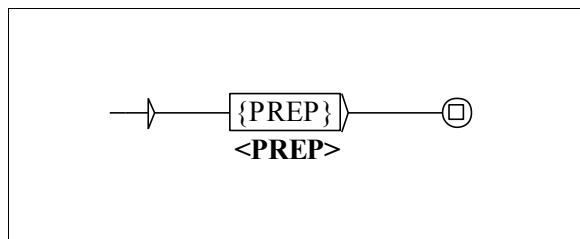
1.6. Un type particulier d'étiquette dans la boîte : le mot est déjà désambiguïsé

Après avoir appliqué une (des) grammaire(s) sur un texte, il est possible de sauvegarder le résultat obtenu et d'appliquer sur ce résultat une ou plusieurs autre(s) grammaire(s). Dans le cadre de l'étiquetage linéaire, ces grammaires travaillent alors sur un texte déjà partiellement étiqueté. Ceci permet de les construire de manière un peu différente, puisqu'elles ont la

⁵ Pour griser un nœud, il suffit de faire précéder le nom de la grammaire qui doit être appelée par le signe du double point (:nom_de_la_grammaire).

⁶Le nœud grisé :1ils représente la grammaire comparable à Graphe3, mais pour le pronom sujet de la troisième personne du pluriel *ils*.

possibilité de se baser sur des mots ou des catégories non ambiguës⁷. Graphe8 illustre le type de notation qu'il faut alors utiliser :



Graphe8

La notation {PREP} indique que le transducteur reconnaît une préposition dont on est certain qu'il s'agit d'une préposition, parce qu'elle a déjà été préalablement analysée comme telle lors d'un premier étiquetage (soit qu'elle a déjà été désambiguïsée, soit qu'elle était non ambiguë). Notons que **la contrainte** doit toujours prendre la forme <PREP> et non {PREP}.

On pourrait penser que {PREP} reconnaît des prépositions non ambiguës (il n'y a pour le mot que la seule hypothèse préposition dans le dictionnaire) même si le texte n'est pas préalablement étiqueté. Or ce n'est pas le cas. Par contre, une grammaire avec ce type de nœud peut tourner sans application préalable de grammaires de levée d'ambiguïtés, mais avec un premier étiquetage qui reconnaît uniquement les mots non ambiguës. Pour ce faire, on tagge sans sélectionner aucune grammaire de désambiguïsation. Seuls les mots non ambiguës sont alors étiquetés (cf. procédure décrite ci-après).

2. Application des grammaires et résultats

2.1. Du lancement du texte à l'application des grammaires

Avant d'appliquer les grammaires de levée d'ambiguïtés, il est nécessaire d'effectuer un certain nombre d'opérations⁸ :

- Text -> open : on ouvre le texte sur lequel on va travailler (il est au format *.txt, dans le répertoire /Corpus).
- À la question « Do you want to preprocess the text ? », répondre Yes.
- Un tableau apparaît, dans lequel il faut effectuer les choix de prétraitement.

Dans l'état actuel, nous ne cochons que les 2 premières cases, et ne sélectionnons aucun dictionnaire ni aucune grammaire par défaut.

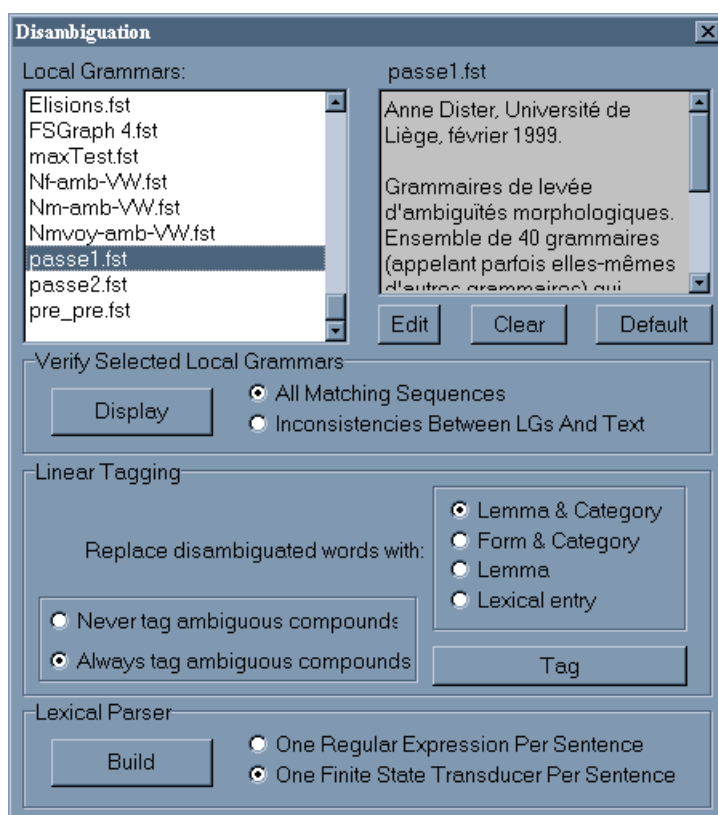
- Cliquer ensuite sur le bouton GO!
- Choisir ensuite Text -> Apply Lexical Ressources.

⁷ Pour le français, nous les appelons « grammaires de passe2 ». (cf. plus loin)

⁸ Nous donnons ici la marche à suivre pour utiliser les grammaires de levée d'ambiguïtés du français disponibles à l'adresse suivante : <http://www.ladl.jussieu.fr/tools/tools.html>.

Pour l'utilisateur qui utilise ses propres grammaires, la procédure est équivalente, moyennant peut-être des choix différents pour le pré-traitement et les dictionnaires.

- Sélectionner pour les mots simples le `delafm.bin` et `Disamb-.dic`, pour les mots composés `DisambComp-.dic` et `DisambSigle-.fst`⁹. Cliquer ensuite sur `Apply dictionaries`.
- Le menu de désambiguïsation est alors actif dans `Text -> Disambiguation`. La fenêtre suivante apparaît :



On peut maintenant appliquer les grammaires sur le texte.

2.2. Application des grammaires

L'idée de base qui a sous-tendu la construction des grammaires du français est d'en faire des grammaires « parfaites » : elles n'ont pas pour ambition de lever la totalité des ambiguïtés d'un texte, mais de fournir des résultats corrects. Nous avons donc préféré garder du *silence* (dans le cas de l'étiquetage, tout le texte n'est pas étiqueté) et éviter le *bruit* (aucune étiquette erronée n'est fournie).

Les grammaires de levée d'ambiguïtés construites pour le français fonctionnent en deux passes : toutes les grammaires de première passe sont regroupées dans le graphe **passe1** (qui fait donc l'union de ces grammaires), et celles de seconde passe dans le graphe **passe2**. Les grammaires de passe2 se basent sur les résultats obtenus après passe1 : elles travaillent sur un texte déjà partiellement désambiguïsé, et comportent donc dans leur chemin des mots ou des catégories déjà désambiguïsés. Certains nœuds des transducteurs de passe2 se présentent comme dans Graphe8 (voir ci-avant).

Pour appliquer des grammaires sur un texte :

⁹ On peut également sélectionner les dictionnaires des toponymes, des sigles, des prénoms, etc.

- Sélectionner passe1 (ou la grammaire - l'union de grammaires - que l'on veut appliquer sur le texte).
- Sélectionner le mode d'étiquetage que l'on désire (voir plus bas) en Linear Tagging
- appuyer sur Tag.
- À la question « Do you want to replace text with new tagged file ? », répondre oui. Intex recharge alors les dictionnaires que l'on avait sélectionnés, et insère dans le texte les étiquettes pour les mots déjà désambiguïsés.
- Retourner dans le menu Text -> Disambiguation et sélectionner passe2 (ou la grammaire - l'union de grammaires - que l'on veut appliquer sur le texte).
- À la question « Do you want to replace text with new tagged file ? », répondre oui.
- Répéter l'opération le nombre de fois qu'on le désire.

2.3. Les résultats de la désambiguïsation : l'étiquetage linéaire

L'une des finalités de la levée d'ambiguïtés est traditionnellement l'étiquetage de textes : celui-ci consiste à associer à chaque mot ou groupe de mots du texte une étiquette, dans laquelle on trouve un certain nombre d'informations (forme, lemme, informations flexionnelles, etc.).

Intex offre également cette possibilité : l'utilisateur a le choix entre 4 types d'étiquettes, plus ou moins complètes :

- Lemma & Category : le lemme et la catégorie grammaticale, sans aucune information de genre ou de nombre pour les noms ou les adjectifs par exemple, ni de conjugaison pour les verbes.
- Form & Category : la forme telle qu'elle est rencontrée dans le texte, et la catégorie grammaticale, sans aucune information de genre ou de nombre, ni de conjugaison.
- Lemma : uniquement le lemme.
- Lexical entry : l'entrée lexicale, telle qu'elle apparaît dans les dictionnaires du système. C'est le type d'étiquette le plus complet, puisqu'il indique non seulement la forme telle qu'on la trouve dans le texte, le lemme mais aussi toutes les informations flexionnelles et éventuellement les traits sémantiques (lorsqu'ils sont présents dans le dictionnaire).

Nous donnons ici un exemple des résultats pour les 4 types d'étiquetage (début du roman d'Agatha Christie, *La maison biscornue*). Nous avons grisé les mots qui sont ou non étiquetés selon le mode d'étiquetage choisi.

a) Lemma & Category

{ce,.PRO} {être,.V} {en,.PREP} {Egypte,.N}, vers la fin de {le,.DET} {guerre,.N}, que {je,.PRO} {faire,.V} {le,.DET} {connaissance,.N} de Sophia Leonides.
 {S} {malgré,.PREP} {son,.DET} {jeunesse,.N}, ({elle,.PRO} {ne,.ADV} {avoir,.V} {que,.ADV} vingt-deux {an,.N} {à,.PREP} {le,.DET} {époque,.N}), {elle,.PRO} {occuper,.V} un poste {assez,.ADV} important {au,.PREPDET} {au,.PREPDET} Foreign {office,.N}.

{S} {je,.PRO}'{avoir,.V} d'abord avec {elle,.PRO} des **relations** professionnelles, **{ce,.PRO}** {qui,.PRO} {me,.PRO} **{permettre,.V}** {très,.ADV} {vite,.ADV} {de,.PREP}'{apprécier,.V} {le,.DET} {compétence,.N} {à,.PREP} laquelle {elle,.PRO} **{devoir,.V}** {son,.DET} {situation,.N} .

b) Form & Category

C'{est,.V} {en,.PREP} {Egypte,.N}, vers la fin de {la,.DET} {guerre,.N}, que {je,.PRO} **{fis,.V}** {la,.DET} {connaissance,.N} de Sophia Leonides.
 {S} {malgré,.PREP} {sa,.DET} {jeunesse,.N}, ({elle,.PRO} {n,.ADV}'{avait,.V} {que,.ADV} **vingt-deux** {ans,.N} {à,.PREP} {l,.DET}'{époque,.N}) , {elle,.PRO} **{occupait,.V}** un poste {assez,.ADV} important {au,.PREPDET} Foreign {office,.N}.
 {S} {j,.PRO}'{eus,.V} d'abord avec {elle,.PRO} des **{relations,.N}** professionnelles, **{ce,.PRO}** {qui,.PRO} {me,.PRO} **{permet,.V}** {très,.ADV} {vite,.ADV} {d,.PREP}'{apprécier,.V} {la,.DET} {compétence,.N} {à,.PREP} laquelle {elle,.PRO} **{devait,.V}** {sa,.DET} {situation,.N} .

c) Lemma

{ce}'{être} {en} {Egypte}, vers la fin {de} {le} {guerre}, {que} {je} **{faire}** {le} {connaissance} {de} Sophia Leonides.
 {S} {malgré} {son} {jeunesse}, ({elle} {ne}'{avoir} {que} **{vingt}-{deux}** {an} {à} {le}'{époque}) , {elle} **{occuper}** {un} poste {assez} important {au} Foreign {office}.
 {S} {je}'{avoir} {de}'{abord} {avec} {elle} des **relations** {professionnel}, **{ce}** {qui} {me} **{permettre}** {très} {vite} {de}'{apprécier} la {compétence} {à} {lequel} {elle} **{devoir}** {son} {situation} .

d) Lexical entry

C'est {en,.PREP} {Egypte,.N+Top:fs}, vers la fin de {la,.le.DET:fs} {guerre,.N:fs}, que {je,.PRO:1s} **fis** {la,.le.DET:fs} {connaissance,.N:fs} de Sophia Leonides.
 {S} {malgré,.PREP} {sa,son.DET:fs} {jeunesse,.N:fs}, ({elle,.PRO:3fs} {n,.ne.ADV}'{avait {que,.ADV} **vingt-deux** {ans,an.N:mp} {à,.PREP} {l,.le.DET:ms:fs}'{époque,.N:fs}) , {elle,.PRO:3fs} **occupait** un poste {assez,.ADV} important {au,.PREPDET:ms} Foreign Office.
 {S} {j,je.PRO:1s}'{eus d'abord avec {elle,.PRO:3fs} des **relations** professionnelles, **ce** {qui,.PRO} {me,.PRO:1s} **permet** {très,.ADV} {vite,.ADV} {d,de.PREP}'{apprécier,.V+t:W} {la,.le.DET:fs} {compétence,.N:fs} {à,.PREP} laquelle {elle,.PRO:3fs} **devait** {sa,son.DET:fs} {situation,.N:fs} .

On remarquera que les résultats ne sont pas équivalents, en terme de pourcentage de mots désambiguïsés, selon le mode d'étiquetage adopté : plus l'étiquette demandée est complète, plus le taux de désambiguïsation est faible. Ainsi, c'est le mode « Lexical entry » qui donne les moins bonnes performances ; à l'opposé, le mode « Lemma » fournit le plus grand nombre de mots étiquetés. L'utilisateur choisit donc le type d'étiquetage selon la complétude de l'information dont il a besoin.

Remarque : les grammaires de passe2 (ou les grammaires construites par l'utilisateur pour tourner sur un texte déjà partiellement désambiguïsé) s'appuient en général sur des catégories grammaticales. Quand on choisit de remplacer les mots du texte uniquement par le lemme (**Mode Lemma**), les étiquettes insérées après passe1 ne contiennent aucune information sur la catégorie grammaticale et ne peuvent alors servir de base à une désambiguïsation ultérieure. Il faut donc, au niveau de passe1, choisir l'étiquetage **Lemma & Category**, et sélectionner le mode Lemma uniquement pour le résultat final (dans notre cas, au moment où l'on applique passe2).

2.3. Verify Selected Local Grammars

Dans le menu Disambiguation se trouve un outil très utile lorsque l'on est amené soi-même à construire des grammaires de levée d'ambiguïtés : il s'agit de la case Verify Selected Local Grammars. Elle permet deux types de vérification, avec :

- a) All Matching Sequences : permet de visualiser dans le texte les séquences couvertes par la/les grammaire(s) sélectionnée(s). On peut également obtenir les résultats sous forme de concordances.
- b) Inconsistencies Between Lgs And Text : repère les incohérences qui existent entre le texte et la/les grammaire(s) sélectionnée(s). Ces incohérences peuvent avoir des origines diverses :
 - l'étiquette utilisée dans la grammaire est incorrecte : AD au lieu de ADV, etc.
 - l'étiquette assignée à un mot dans la grammaire est incohérente avec celle présente dans le dictionnaire. Par exemple, dans la version actuelle d'Intex, *ne* n'est pas étiqueté comme adverbe et nous utilisons cette contrainte dans plusieurs grammaires¹⁰.
 - une grammaire de passe2 peut contredire des résultats fournis en passe1.

2.4. Les mots composés : Never tag ambiguous compounds / Always tag ambiguous compounds

Dans la case Linear Tagging se trouve une alternative dont nous n'avons pas encore parlé :

- Never tag ambiguous compounds
- Always tag ambiguous compounds

Il s'agit soit de ne jamais étiqueter les composés ambigus (ils sont alors analysés comme des séquences libres), soit de toujours les étiqueter (on les considère d'office comme non ambigus) : quel que soit le choix qui est fait, le résultat fourni est faussé. Et dès lors que l'on connaît le nombre important de mots composés présents dans un texte, on s'aperçoit aisément de **l'impasse à laquelle conduit l'étiquetage linéaire.**

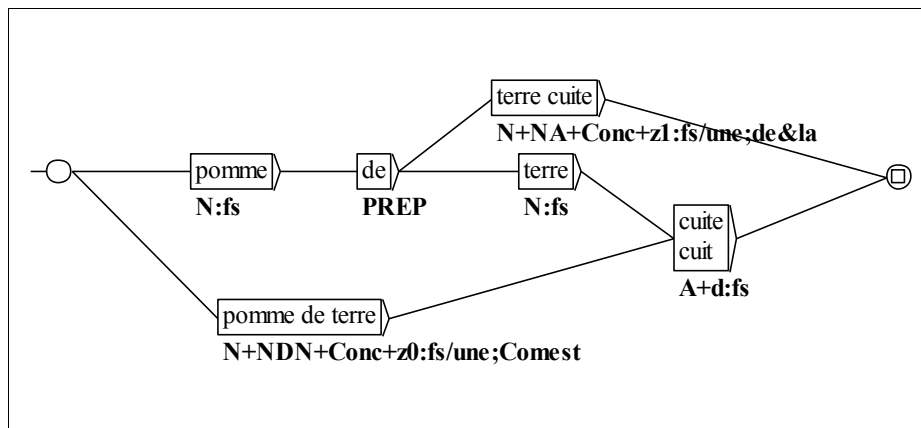
La seule solution envisageable est alors de ne plus obtenir les résultats de la désambiguïsation sous la forme traditionnelle d'un texte linéaire mais sous la forme d'un transducteur.

2.5. Lexical Parser

Ce menu permet d'obtenir les résultats de la levée d'ambiguïtés sous forme d'expression régulière ou encore, et c'est ce qui nous intéresse ici, sous la forme d'un transducteur.

Ainsi, une séquence ambiguë comme « la pomme de terre cuite » se présentera sous la forme suivante :

¹⁰ L'une des raisons d'être du dictionnaire Disamb-.dic est de rétablir des hypothèses omises dans le delafm.

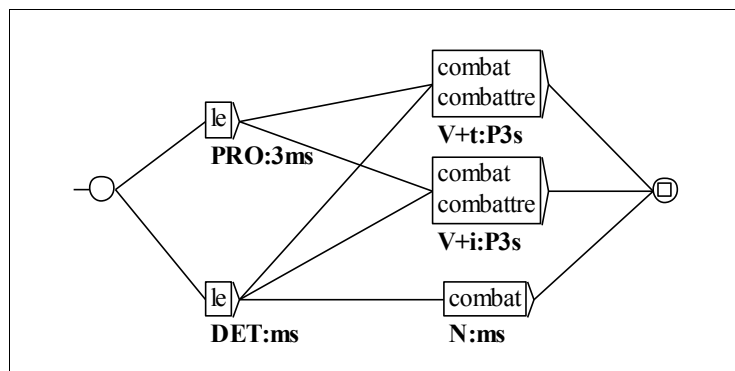


Graphe9

On ne choisit plus arbitrairement s'il s'agit d'une séquence figée ou libre : si l'on n'a pu résoudre l'ambiguïté, toutes les solutions sont proposées dans des chemins parallèles. Dans le cas où l'ambiguïté est résolue, un seul chemin subsiste.

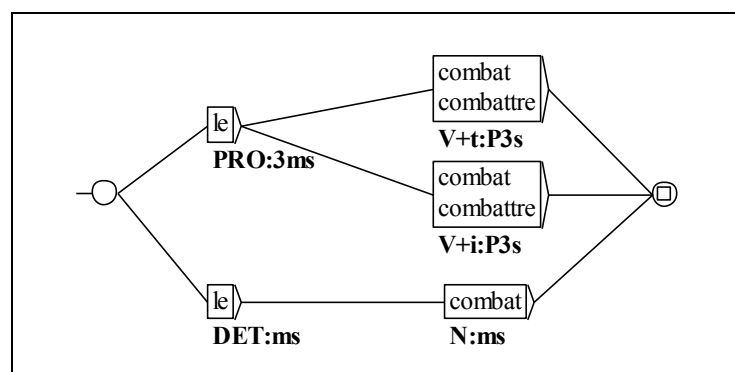
Ce type de représentation a encore un autre avantage : on se souvient que Graphe5 ne donnait aucun résultat lors de l'étiquetage linéaire de la séquence « le combat ». Dès lors que l'on travaille directement sur l'automate du texte, on voit l'efficacité d'une telle grammaire.

Graphe10 présente le transducteur du texte sans utilisation de Graphe5 :



Graphe10

Tandis qu'avec application de Graphe5 sur l'automate du texte, nous obtenons Graphe11 :



Graphe11

On le voit, même si aucune solution n'a été choisie, la grammaire a néanmoins permis d'éliminer certains chemins. Dans de nombreux cas, des suppressions successives de transitions permettent de lever un certain nombre d'ambiguïtés.

Conclusion

On l'a vu, la seule solution satisfaisante du point de vue théorique pour la levée d'ambiguïtés est d'abandonner l'étiquetage linéaire au profit d'un traitement de l'automate du texte. Seul celui-ci peut représenter des résultats parfaits qui pourront servir de base à des traitements ultérieurs sur le texte, en particulier l'analyse syntaxique globale de la phrase.

La prochaine version d'Intex (4.20) privilégiera donc ce mode de désambiguïsation : l'utilisateur pourra appliquer en cascade des grammaires à l'automate du texte en sauvant chaque résultat intermédiaire¹¹.

Anne Dister
CIPL, Université de Liège (Bât. A4)
1b Quai Roosevelt
4000 Liège
Belgique
adister@ulg.ac.be

RÉFÉRENCES

- Dister, Anne. 1999. La levée d'ambiguïtés morphologiques par grammaires locales. Le système Intex . In *Actes des XIIIes Journées de Linguistique*, (25 et 26 mars 1999), Québec. A paraître.
- Dister, Anne. 1999. De l'étiquetage traditionnel au transducteur du texte : Intex et la levée d'ambiguïtés par grammaires locales. *RISSH, Revue Informatique et Statistiques dans les Sciences Humaines*, t. , n° 1-4.
- Laporte, Éric. 1995. Levée d'ambiguïtés par grammaires locales. *Lingvisticae Investigationes Supplementa*, Amsterdam/Philadelphia: John Benjamins.
- Silberztein, Max. 1993. *Dictionnaires électroniques et analyse automatique de textes, Le système INTEX*. Paris: Masson.
- Silberztein, Max. 1996. Expérience d'étiquetage avec levée d'ambiguïtés. In *Actes des Premières Journées INTEX (21-22 mars 1996)*, Paris: LADL, Université Paris 7, CNRS.
- Silberztein, Max. 1997. The lexical analysis of natural languages . In *Finite-state language processing*, E. Roche, Yves Schabès (eds.), Cambridge, Mass./London: MIT Press.

¹¹ En l'état actuel, il a déjà la possibilité de le faire, mais en travaillant directement en ligne de commande :
programme etiq pour construire le FST du texte ;
programme interg pour construire l'intersection du FST du texte et du FST de la grammaire de levée d'ambiguïtés.

