



La correction complète de ce tutoriel est disponible sur :

Préparation

La première chose à faire est d'installer le logiciel libre Unitex. Il faut télécharger la version 3.2 ou une version ultérieure : <https://unitexgramlab.org/>. En cas de difficultés d'installation, le manuel est aussi accessible en ligne : <https://unitexgramlab.org/releases/3.2/man/Unitex-GramLab-3.2-usermanual-fr.pdf>. Il faut alors consulter les sections 1.3-1.7.

Une fois le logiciel installé, le manuel Unitex est disponible via le menu *Help/Manuals*.

Ouvrir un texte : voir le manuel, section 2.4-2.5.5.

Créer un graphe : voir le manuel, section 5.2.

En bref, pour créer un graphe, il faut retenir les trois points suivants :

1. Créer (ou supprimer) un chemin entre deux boîtes : cliquer sur la première, puis sur la deuxième boîte.
2. Créer une boîte : clic-droit de la souris et choisir *Create box*. Si une boîte est sélectionnée, un chemin sera automatiquement tracé entre la boîte sélectionnée et la nouvelle boîte.
3. Remplir une boîte : écrire sur la barre de formule et valider. Si la barre de formule est totalement vide, la boîte est supprimée. Le symbole $\langle E \rangle$ désigne l'élément vide (*empty*) et permet de créer une boîte vide.

Préparation

Le plus simple est de télécharger le fichier :

https://tln.lifat.univ-tours.fr/medias/fichier/tutoriel-priseenmain-unitex-creationdictionnaires-denis-maurel_1573493543126-zip?ID_FICHE=334600&INLINE=FALSE

et de le dézipper dans votre dossier personnel Unitex (les fichiers se placeront au bon endroit).

Puis de passer à l'ouverture du corpus.

En cas d'impossibilité, poursuivre ci-dessous.

Création de dossiers

Plaçons-nous dans notre dossier personnel Unitex, dans *French*.

1. Dans *Corpus* : Créons un dossier nommé *80jours* et glissons dans ce dossier le fichier *80jours.txt* ("Le tour du monde en 80 jours" de Jules Verne), distribué avec Unitex ;
2. Dans *Inflexion* : Créons un dossier pour l'ensemble du tutoriel, nommé *Tutoriel_Unitex_Dicos*.
3. Dans le dossier *Inflexion*, copions les deux fichiers *Equivalences.txt* et *Morphology.txt*, puis collons-les dans le dossier *Inflexion/Tutoriel_Unitex_Dicos*.

Ouverture du corpus

Commençons par ouvrir, dans le menu *Text*, le fichier *French/Corpus/80jours/80jours.txt*, en répondant *No* à la question *Do you want to preprocess the text*.

Création du fichier dictionnaire des formes vedettes

Ouvrons le menu *File Edition/New File* et enregistrons le fichier dans le répertoire *Dela* sous le nom *MonDico.dic*. Ce fichier sera notre dictionnaire de formes vedettes. Nous allons distinguer la flexion des mots monolexicaux et celle des mots polylexicaux¹.

1 Flexion des mots monolexicaux

Voir le manuel, section 3.5.1.

1.1 Flexion par simple suffixation

1.1.1 Les noms masculins qui prennent un s au pluriel

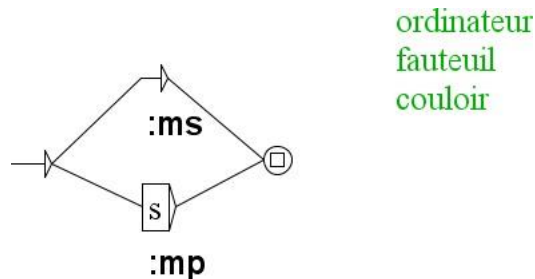
Tapons dans le fichier *MonDico.dic* les trois lignes suivantes (sans espaces et en tapant *Entrée* à la fin de la troisième ligne) :

ordinateur,N100 fauteuil,N100 couloir,N100
--

¹ Il est aussi possible de créer des règles de flexion qui prennent en compte la racine, pour les langues sémitiques. Voir le manuel, section 3.5.4.

N désigne la catégorie grammaticale, nom, et *N100*² est le nom du graphe de flexion associé, graphe que nous allons créer et enregistrer (comme les suivants) dans le dossier *French/Inflexion/Tutoriel_Unitex_Dicos*. Enregistrons et fermons le fichier *MonDico.dic*.

Le graphe *N100* va permettre de fléchir les noms masculins qui prennent un *s* au pluriel. Nous allons utiliser (mais ce n'est pas obligatoire) les codes du dictionnaire fourni dans la distribution d'Unitex, *Dela_fr.bin* : *ms* pour le masculin singulier et *mp* pour le masculin pluriel. Pour mémoire, nous ajoutons un commentaire avec les exemples concernés (une boîte commençant par un */*).

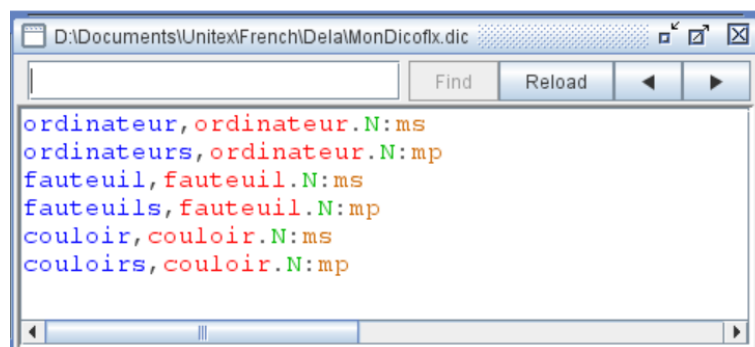


Voir la vidéo : https://tln.lifat.univ-tours.fr/medias/video/n100_1580218891978-mp4?ID_FICHE=334600&INLINE=FALSE.

Ce graphe doit être compilé : cliquons sur le bouton *Compile graph*. Une erreur apparaît : *Main graph matches epsilon! Error: the main graph N100 recognize <E>*. Cette erreur ne nous concerne pas, elle signifie que ce graphe ne peut être utilisé en recherche d'occurrences dans le menu *Locate pattern*. Cliquons sur *OK*.

Lançons le programme de flexion. Ce programme va créer un dictionnaire appelé *MonDicoflx.dic* (par suffixation de *flx* sur le nom du fichier). Pour cela, dans le menu *DELA* (et non cette fois-ci dans le menu *File Edition*), ouvrir le fichier *MonDico.dic*. Ici, il n'est plus question de modifier le fichier, mais d'utiliser les programmes concernant les dictionnaires³.

Fléchissons le dictionnaire *MonDico.dic* en cliquant sur le menu *DELA/Inflect*. Cliquons sur le bouton *SET*, sur le dossier *French/Inflexion/Tutoriel_Unitex_Dicos* et sur le bouton *Open*. Cliquons sur le bouton *Inflect Dictionary*. Le fichier *MonDicoflx.dic* est créé et s'affiche à l'écran.



Nous pouvons trier ce dictionnaire⁴ en cliquant sur le menu *DELA/Sort Dictionary*.

² Nous commençons la numérotation à 100 pour éviter la confusion avec les graphes de la distribution d'Unitex.

³ Si vous constatez une erreur dans l'écriture de votre fichier, il faut le fermer dans le menu *DELA* et le rouvrir dans le menu *File Edition*.

⁴ Attention que l'ancienne version est remplacée par la nouvelle.

Si nous souhaitons utiliser ce dictionnaire pour analyser le texte *80jours*, il faut cliquer sur le menu *DELA/Compress into FST* et sur le bouton *OK*.

Remarque : Cette commande crée deux fichiers, un fichier *MonDicoFlx.bin*⁵ et un fichier *MonDicoFlx.inf*⁶. Si vous souhaitez copier votre dictionnaire sur un autre ordinateur, les deux sont nécessaires.

Fermons les deux fenêtres des dictionnaires *MonDico.dic* et *MonDicoFlx.dic*. Ouvrons maintenant le menu *Text/ApplyLexical Resources* et cliquons sur le bouton *Clear*. Choisissons (à gauche) le fichier *MonDicoFlx.bin* et cliquons sur le bouton *Apply*. La *Word List* s'ouvre. Deux mots de notre dictionnaire figurent dans le roman de Jules Verne : *couloirs* et *fauteuils*. Les 9 429 autres mots du roman apparaissent dans la fenêtre des mots inconnus.

Fermons la *Word List*.

1.1.2 Plusieurs autres exemples

Ouvrons à nouveau le fichier *MonDico.dic* par le menu *File Edition/Open.../Dictionaries* et répondons *OK* à l'avertissement *This is not necessarily the text being processed by Unitex* qui signifie que ce fichier ne peut servir à l'analyse textuelle⁷. Ajoutons à notre dictionnaire de formes vedettes plusieurs autres exemples :

- *chaise, table, porte* qui sont des noms féminins qui prennent un *s* au pluriel. Graphe N101.
- *bateau, bureau, tableau* qui sont des noms masculins qui prennent un *x* au pluriel. Graphe N102.
- *Blanc* qui est un nom masculin qui ajoute *he* au féminin et qui prend un *s* au pluriel. Graphe N103. De même pour les adjectifs *franc, blanc*. Graphe A103.
- *ami, voisin, invité* qui sont des noms masculins qui prennent un *s* au pluriel et admettent un féminin qui prend un *e* au singulier et un *es* au pluriel. Graphe N104. De même pour les adjectifs *grand, fort, intelligent*. Graphe A104.
- *Français, bourgeois, mauvais* qui sont des noms masculins singuliers et pluriels qui admettent un féminin qui prend un *e* au singulier et un *es* au pluriel. Graphe N105. De même pour les adjectifs *français, bourgeois, mauvais*. Graphe A105.
- *collègue, arbitre, artiste* qui ont des noms masculins et féminins singuliers qui prennent un *s* au pluriel. Graphe N106. De même pour les adjectifs *athlétique, balnéaire, crédible*. Graphe A106.

Rappelons le modèle à suivre (sans espaces et en tapant *Entrée* à la fin de la dernière ligne) :

chaise,N101 table,N101 ... crédible,A106

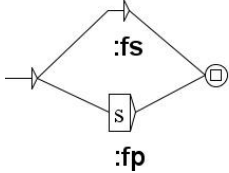
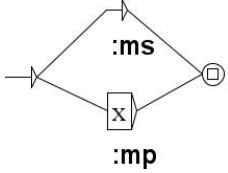
Enregistrons et fermons le fichier *MonDico.dic*.

⁵ Ce premier fichier contient les mots. Il est binaire et ne peut donc pas être lu par un traitement de texte.

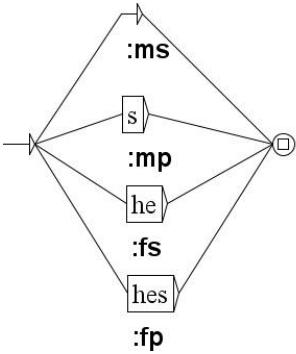
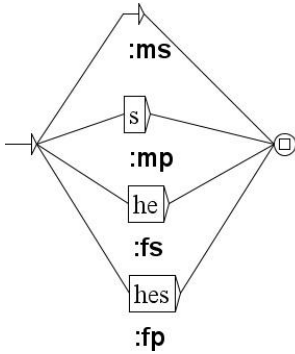
⁶ Ce deuxième fichier contient la factorisation des étiquettes associées aux mots. Il est lisible.

⁷ Sauf en le fermant et en l'ouvrant dans le menu *Text/Open*.

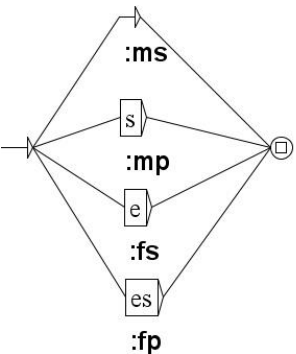
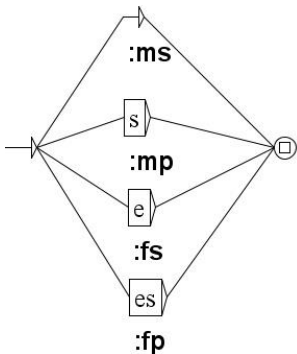
Créons (et compilons) les graphes N101 et N102 : ces graphes se ressemblent beaucoup, nous pouvons utiliser le menu *FSGraphs/Save as...* et changer quelques détails dans les boîtes, sans oublier les exemples. N'oublions pas que tous les graphes doivent être compilés.

 chaise table porte	 bateau bureau tableau
N101	N102

Créons (et compilons) les graphes et N103 et A103. Nous pouvons repartir du graphe N100 et créer deux nouvelles boîtes. Le graphe A103 est identique au graphe N103⁸, aux exemples près.

 Blanc	 franc blanc
N103	A103

Créons les graphes N104, A104, N105, A105, N106 et A106 :

 ami voisin invité	 grand fort intelligent
N104	A104

⁸ C'est une particularité du français d'avoir des formes et des flexions identiques pour les noms et les adjectifs à quelques exceptions près.

N105	A105
N106	A106

Nous pouvons maintenant ouvrir le dictionnaire *MonDico.dic* par le menu *DELA/Open Recent* et le fléchir par le menu *DELA/Inflect*. Le nouveau dictionnaire *MonDicoFlx.dic* s'ouvre avec 102 lignes. Vérifions les nouvelles flexions⁹ et, si tout va bien, trions-le par le menu *DELA/Sort Dictionary*. Puis rendons-le opérationnel par le menu *DELA/Compress into FST*. Fermons les deux fenêtres des dictionnaires *MonDico.dic* et *MonDicoFlx.dic*. Testons ce nouveau dictionnaire sur notre fichier *80jours* (ne pas oublier de cliquer sur le bouton *Clear* dans le menu *Text/ApplyLexical Resources*). Nous obtenons 59 entrées. Fermons la *Word List*.

1.2 L'opérateur L

Pour suffixer un mot à partir d'une autre lettre que la dernière, nous utilisons l'opérateur *L* (pour *Left*) pour nous déplacer de droite à gauche à partir de la fin du mot.

1.2.1 Flexion du mot *cheval*

Les mots *cheval*, *journal*, *cristal* sont des noms masculins qui remplacent le *l* final par *ux* au pluriel. À partir de la fin du mot, on se déplace donc d'une lettre à gauche (opérateur *L*) et on ajoute *ux*. Cette règle se note *Lux* dans la boîte¹⁰.

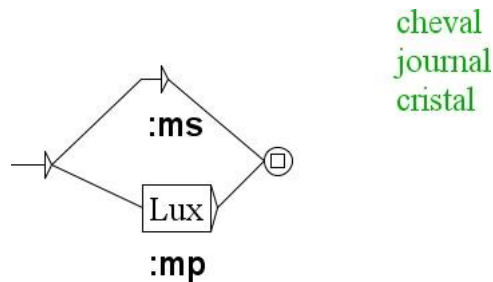
Ajoutons trois lignes dans le fichier *MonDico.dic* par le menu *File Edition/Open.../Dictionaries* et répondons *OK*.

```
cheval,N107
journal,N107
cristal,N107
```

Enregistrons et fermons ce fichier. Créons donc le graphe Graphe *N107*, par exemple à partir du graphe *N100* et du menu *Save as...*

⁹ Remarque : nous pouvons fléchir ce dictionnaire dès la construction du premier graphe. Nous obtiendrons une erreur pour les graphes n'existant pas et un dictionnaire des formes fléchies correspondant aux graphes existants.

¹⁰ Les flexions étant écrites en minuscules, les majuscules sont réservées pour les opérateurs.



Puis ouvrons le fichier *MonDico.dic* par le menu *DELA/Open Recent* et fléchissons-le par le menu *DELA/Inflect*. Le nouveau dictionnaire *MonDicoflx.dic* s'ouvre avec 108 lignes. Vérifions les nouvelles flexions et, si tout va bien, trions-le par le menu *DELA/Sort Dictionary*. Créons, par le menu *DELA/Compress into FST*, un fichier utilisable. Fermons les deux fenêtres des dictionnaires *MonDico.dic* et *MonDicoflx.dic*. Ouvrons le menu *Text/ApplyLexical Resources* (sans oublier de cliquer sur le bouton *Clear*). Nos six nouveaux mots se trouvent dans le texte, notre liste de mots reconnus passe à 65 entrées. Fermons la *Word List*.

1.2.2 Plusieurs autres exemples

Ouvrons à nouveau le fichier *MonDico.dic* dans le menu *File Edition/Open.../Dictionaries*, OK. Ajoutons à notre dictionnaire de formes vedettes plusieurs autres exemples utilisant une ou plusieurs fois l'opérateur *L* :

- *boulangier, cordonnier, cuisinier* qui sont des noms masculins qui remplacent le *er* par *ère* au féminin et qui prennent un *s* au pluriel. Graphe *N108*. De même pour les adjectifs *dernier, étranger, fier*. Graphe *A108*.
- *vendeur, chanteur, défenseur* qui sont des noms masculins qui remplacent le *r* par *se* au féminin et qui prennent un *s* au pluriel. Graphe *N109*.
- *acteur, lecteur, éditeur* qui sont des noms masculins qui remplacent le *eur* par *rice* au féminin et qui prennent un *s* au pluriel. Graphe *N110*.
- *chanter, aimer, danser* sont des verbes qui éliminent le *r* ou le *er* au présent de l'indicatif¹¹. Graphe *V100*.
- *manger, nager, plonger* sont des verbes qui éliminent le *r* au présent de l'indicatif. Graphe *V101*.
- *rapiécer, agacer, bercer* sont des verbes qui éliminent le *r* ou le *er* en remplaçant le *c* par un *ç* au présent de l'indicatif. Graphe *V102*.

Ajoutons donc les lignes (en tapant *Entrée* à la fin de la dernière ligne) :

```
boulangier,N108
...
bercer,V102
```

Fermons le fichier *MonDico.dic*. Créons (et compilons) les graphes *N108*, *A108*, *N109* et *N110* :

¹¹ Dans les flexions verbales, nous nous limiterons ici au présent de l'indicatif. La distribution d'Unitex contient quelques flexions verbales complètes qu'il est possible de consulter.

N108	A108
N109	N110

Puis les graphes de verbes¹² :

V100	V101	V102

Il nous reste à ouvrir le fichier *MonDico.dic* par le menu *DELA/Open Recent* et à le fléchir par le menu *DELA/Inflect*. Le nouveau dictionnaire *MonDicoFlx.dic* comprend 210 lignes. Après un tri (menu *DELA/Sort Dictionary*) et une transformation (menu *DELA/Compress into FST*), fermons les deux fenêtres *MonDico.dic* et *MonDicoFlx.dic*. Ouvrons le menu *Text/ApplyLexical*

¹² Pour les verbes aussi, sans que cela soit obligatoire, nous utilisons les codes de la distribution d'Unitex : V pour verbe, P pour présent et 1, 2, 3 pour les personnes.

Resources (sans oublier de cliquer sur le bouton *Clear*). La *Word List* recense 88 mots reconnus. Fermons la *Word List*.

1.3 Les opérateur *R* et *C*

Pour se déplacer d'une lettre vers la droite, nous utilisons l'opérateur *R* (*Right*) qui, combiné à l'opérateur *L*, nous permet de conserver certaines lettres du mot vedette.

Quel point commun entre les verbes *peler*, *gérer* et *acheter* ? Ce sont des verbes qui parfois remplacent le *e* ou le *é* par un *è* (en conservant la consonne qui suit). En effet, la première personne du présent de l'indicatif peut s'obtenir en enlevant quatre lettres (*LLLL*), en écrivant un *è*, puis en se déplaçant vers la droite (*R*) et en ajoutant un *e* (graphe V103).

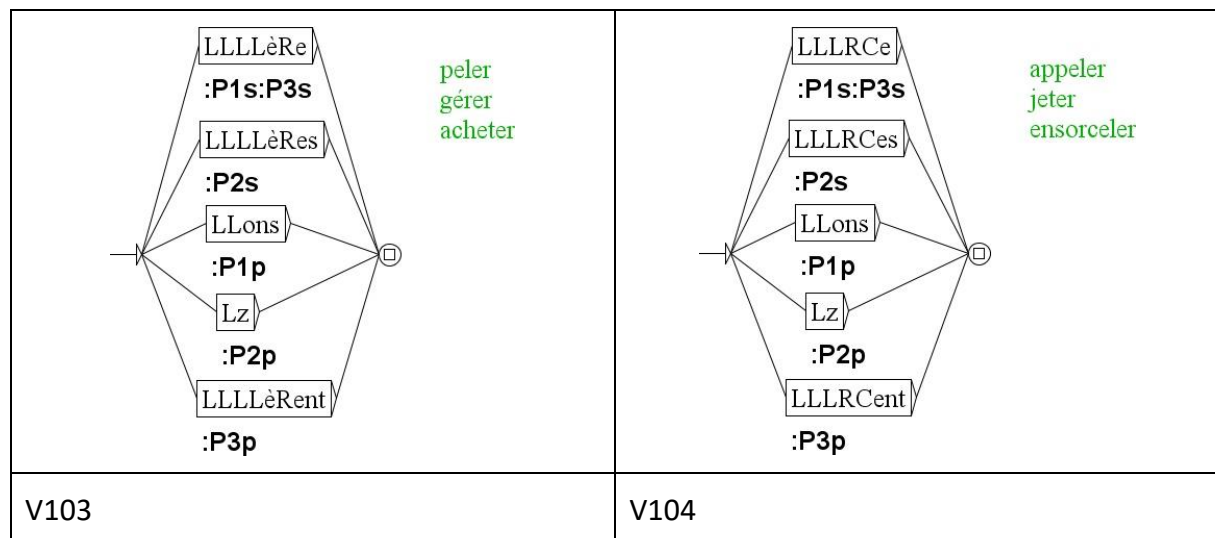
Quant à l'opérateur *C*, il duplique la lettre en cours (c'est-à-dire située à gauche du curseur), ce qui permet par exemple de fléchir par un seul graphe les verbes *appeler*, *jeter*, *ensorceler* qui dupliquent parfois la consonne située avant le *er*. Ainsi, pour obtenir la première personne du présent de l'indicatif, il faut reculer de trois lettres (*LLL*), revenir à droite et dupliquer la lettre qu'on vient de passer (*RC*), puis ajouter un *e* (graphe V104).

Cet opérateur est aussi utile pour quelques noms et adjectifs comme *colonel*¹³, *champion* (graphe N111) et *bon*, *ancien*, *nul* (graphe A111).

Ajoutons donc les lignes (en tapant *Entrée* à la fin de la dernière ligne) :

```
peler,V103
...
ensorceler,V104
colonel,N111
...
nul,A111
```

Fermons le fichier *MonDico.dic*. Créons (et compilons) les graphes V103 et V104, puis N111 et A111 :



¹³ Donc, en dernière position, l'opérateur *C* permet de dupliquer la dernière lettre pour commencer un suffixe, comme pour le mot *colonel* qui duplique le *l* au féminin.

N111	A111

Après avoir ouvert le fichier *MonDico.dic* par le menu *DELA/Open Recent* et l'avoir fléchi par le menu *DELA/Inflect*, le fichier *MonDicoFlx.dic* possède 266 lignes. Nous le trions (menu *DELA/Sort Dictionary*) et le transformons (menu *DELA/Compress into FST*). Puis nous fermons les deux fenêtres *MonDico.dic* et *MonDicoFlx.dic* et ouvrons le menu *Text/ApplyLexical Resources*. La *Word List* recense 109 mots reconnus. Nous la fermons.

Remarque : il existe d'autres opérateurs décrits dans le manuel Unitex.

2 Flexion des mots polylexicaux

Voir le manuel, section 11.3.2.

Pour traiter la flexion des mots polylexicaux, nous allons créer des graphes ayant autant de boîtes que de *tokens* pour la description, plus une boîte pour le résultat. Rappelons que pour Unitex, les unités de traitement, ou *tokens*, sont les séquences de lettres¹⁴ ou tous les autres caractères pris isolément.

2.1 Exemple du mot *bateau-pilote*

Les mots *bateau-pilote*, *bateau-mouche*, *carte-mère* sont des mots polylexicaux (composés de trois tokens) dont le genre est celui du premier token. Nous allons les fléchir par un graphe que nous appellerons *NC_XX1*¹⁵.

Ajoutons trois lignes dans le fichier *MonDico.dic* par le menu *File Edition/Open.../Dictionaries* et répondons *OK*. Ces lignes contiennent à la fois des informations sur le graphe de flexion global et, entre parenthèses, mais sans espaces, des informations sur les graphes de flexion particuliers à chaque mot-token : mot vedette, graphe de flexion et information flexionnelle.

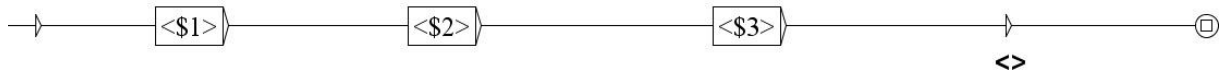
```
bateau(bateau.N102:ms)-pilote(pilote.N100:ms),NC_XX1
bateau(bateau.N102:ms)-mouche(mouche.N101:fs),NC_XX1
carte(carte.N101:fs)-mère(mère.N101:fs),NC_XX1
```

¹⁴ Les lettres sont, pour une langue donnée, définies par le fichier *alphabet.txt*.

¹⁵ Le préfixe *NC_* indique qu'il s'agit de la flexion d'un nom composé.

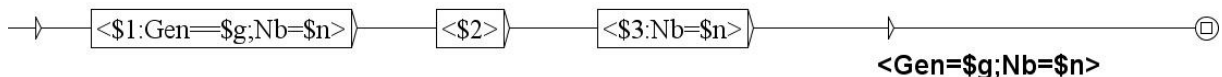
Le graphe NC_XX1 est composé de quatre boites (pour les trois tokens et le résultat) :

bateau-pilote
bateau-mouche
carte-mère



Le deuxième token est un trait d'union, donc sans flexion. Nous allons utiliser les mots-clés¹⁶ *Gen* pour genre et *Nb* pour nombre. La formule *Nb=\$n* permet d'associer la variable *\$n* au nombre des premier et troisième tokens. La formule *Gen==\$g* associe la variable *\$g* au nombre du premier token et ajoute que le genre du mot polylexical est hérité¹⁷ de celui du premier token.

bateau-pilote
bateau-mouche
carte-mère



Ce graphe indique précisément dans la quatrième boite :

1. Que le genre du mot polylexical est celui du premier token et que le genre du troisième token n'est pas pris en compte.
2. Que le nombre du mot polylexical est celui des premier et troisième tokens, identiques entre eux et variables (singulier et pluriel).

Ainsi, nous obtiendrons deux lignes dans le dictionnaire *MonDicoflx.dic* avec les mots *bateau-pilote* et *bateaux-pilotes* et nous éviterons les entrées erronées, *bateaux-pilote* et *bateau-pilotes*.

Fermons ce graphe après l'avoir compilé. Ouvrons le fichier *MonDico.dic* par le menu *DELA/Open Recent* et fléchissons-le par le menu *DELA/Inflect*. Le nouveau dictionnaire *MonDicoflx.dic* comprend 272 lignes. Après un tri (menu *DELA/Sort Dictionary*) et une transformation (menu *DELA/Compress into FST*), fermons les deux fenêtres *MonDico.dic* et *MonDicoflx.dic*. Ouvrons le menu *Text/ApplyLexical Resources* (sans oublier de cliquer sur le bouton *Clear*). La *Word List* recense 111 mots reconnus¹⁸. Fermons la *Word List*.

¹⁶ Ces mots-clés sont définis dans le fichier *Morphology.txt*, qui doit être présent là où se trouvent les graphes de flexion.

¹⁷ C'est-à-dire que *bateau-pilote* et *bateau-mouche* sont masculin comme *bateau* et que *carte-mère* est féminin comme *carte*.

¹⁸ En fait, cent neuf *simple word lexical entries* et deux *compound lexical entries*.

2.2 Plusieurs autres exemples

Ouvrons une dernière fois le fichier *MonDico.dic* par le menu *File Edition/Open.../Dictionaries*, OK. Ajoutons à notre dictionnaire de formes vedettes plusieurs autres exemples de mots polylexicaux :

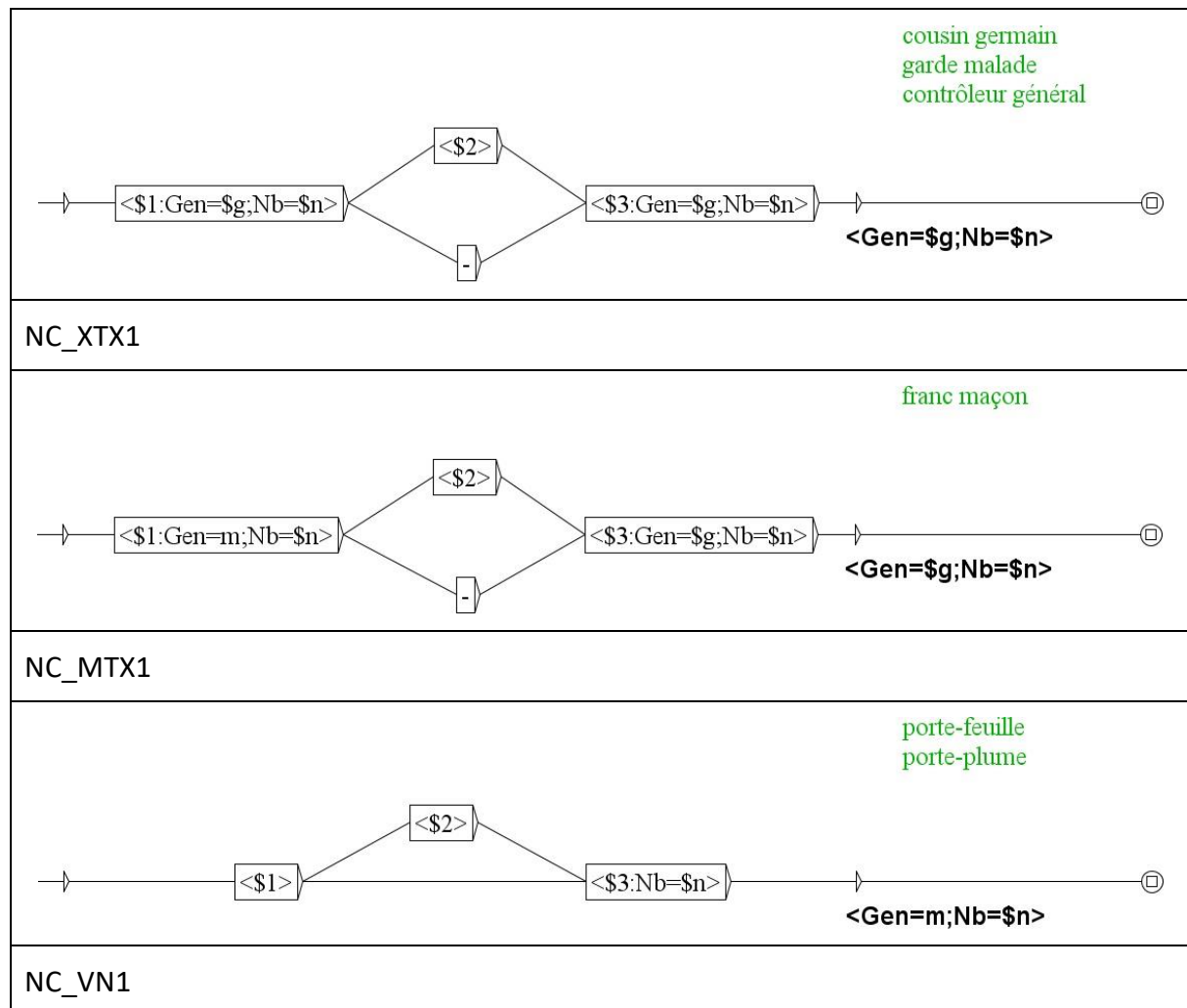
- *cousin germain, garde malade, contrôleur adjoint* qui sont des mots polylexicaux composés de trois tokens. Le genre et le nombre du premier et du troisième tokens sont identiques. Un tiret peut être ajouté¹⁹. Graphe *NC_XX1*.
- *franc maçon* qui est un mot polylexicaux composé de trois tokens. Le nombre du premier et du troisième tokens sont identiques. Le genre du premier token reste masculin. Un tiret peut être ajouté. Graphe *NC_MTX1*.
- *porte-feuille, porte-plume* qui sont des mots polylexicaux composés de trois tokens. Le premier token est invariable. Le premier et le troisième tokens peuvent être soudés. Graphe *NC_VN1*.
- *bonhomme, gentilhomme* sont des mots polylexicaux composés d'un seul token, mais un accord en nombre est quand même réalisé. En fait, nous allons dans le dictionnaire les séparer en deux tokens (sans espace, bien sûr, les séparateurs seront les parenthèses). Graphe *NC_XX2*.
- *cordons bleu, arme blanche, heure matinale* qui sont des mots polylexicaux composés de trois tokens, dont le genre est celui du premier token. Le genre et le nombre du premier et du troisième tokens sont identiques. Le mot vedette associé au troisième token n'est pas le mot écrit qui en est une flexion. Graphe *NC_XX3*.
- *porte-serviette* qui est un mot polylexicaux composé de trois tokens avec deux singuliers. Le premier token est invariable. Le premier et le troisième tokens peuvent être soudés. Graphe *NC_VN2*.
- *auteur compositeur interprète, boucher charcutier traiteur* sont des mots polylexicaux (composé de cinq tokens). Le genre et le nombre du premier, du troisième et du cinquième tokens sont identiques. Des tirets peuvent être ajoutés, mais obligatoirement entre chaque mot. Graphe *NC_XXTX1*.

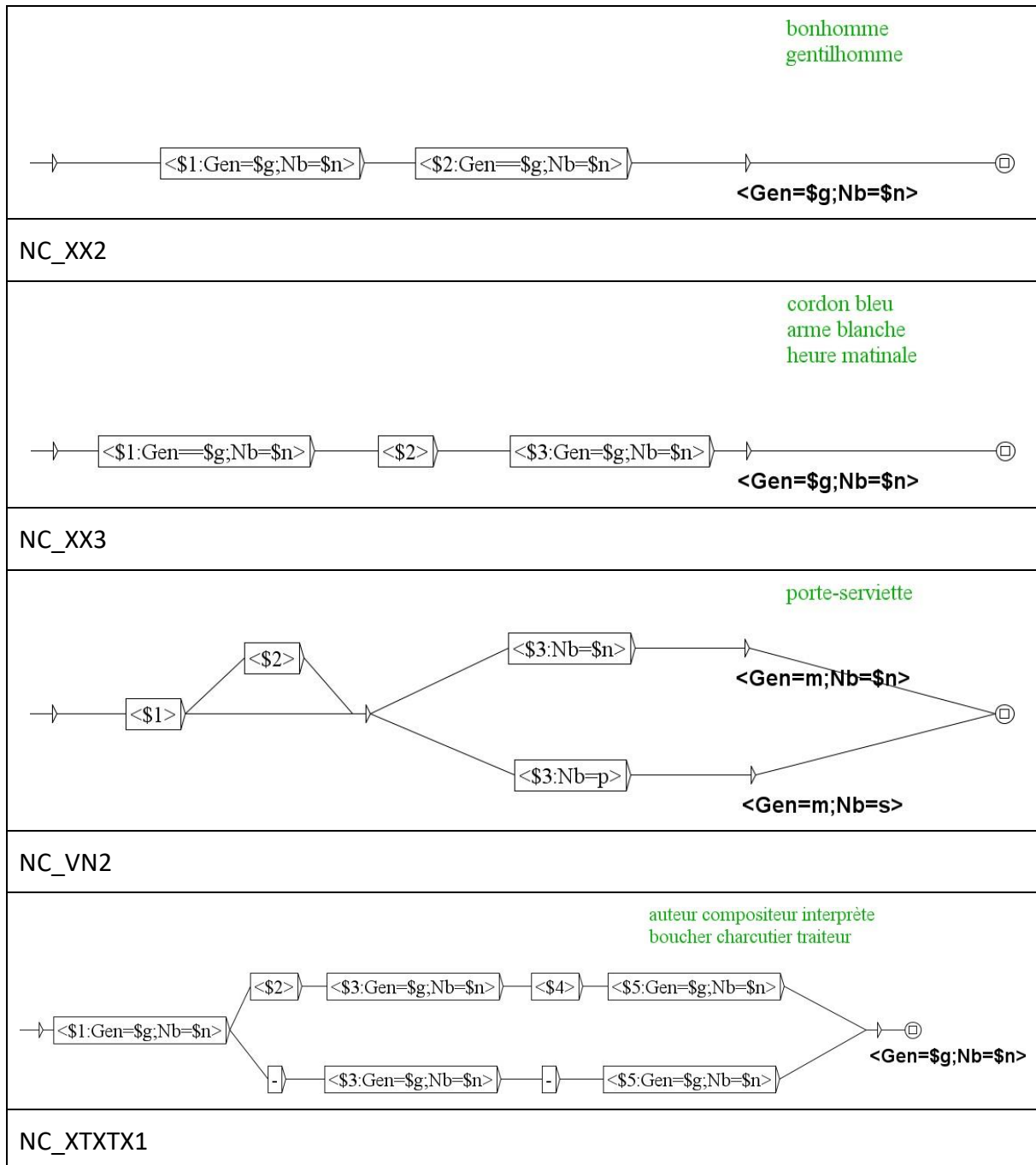
¹⁹ La gestion des espaces n'étant pas simple dans Unitex, on privilégiera l'écriture de l'espace dans le dictionnaire à fléchir et la possibilité du tiret sera notée sur le graphe de flexion.

On obtient le dictionnaire final *MonDico.dic* par l'ajout des quatorze lignes suivantes :

```
cousin(cousin.N104:ms) germain(germain.A104:ms),NC_XTX1
garde(garde.N106:ms) malade(malade.N106:ms),NC_XTX1
contrôleur(contrôleur.N109:ms) adjoint(adjoint.A104:ms),NC_XTX1
franc(franc.A103:ms) maçon(maçon.N104:ms),NC_MTX1
porte-feuille(feuille.N101:fs),NC_VN1
porte-plume(plume.N101:fs),NC_VN1
bon(bon.A111:ms)homme(homme.N100:ms),NC_XX2
gentil(gentil.A111:ms)homme(homme.N100:ms),NC_XX2
cordon(cordon.N100:ms) bleu(bleu.A104:ms),NC_XX3
arme(arme.N101:fs) blanche(blanche.A103:fs),NC_XX3
heure(heure.N101:fs) matinale(matinal.A104:fs),NC_XX3
porte-serviette(serviette.N101:fs),NC_VN2
auteur(auteur.N110:ms) compositeur(compositeur.N110:ms) interprète(interprète.N106:ms),NC_XTXX1
boucher(boucher.N108:ms) charcutier(charcutier.N108:ms) traiteur(traiteur.N109:ms),NC_XTXX1
```

Construisons (et compilons) les sept graphes correspondants :





Fermons tous les graphes par le menu *Graphs/Close all*. Ouvrons le fichier *MonDico.dic* par le menu *DELA/Open Recent* et fléchissons-le par le menu *DELA/Inflect*. Le dictionnaire final *MonDicoFlx.dic* comprend 344 lignes. Après un tri (menu *DELA/Sort Dictionary*) et une transformation (menu *DELA/Compress into FST*), fermons les deux fenêtres *MonDico.dic* et *MonDicoFlx.dic*. Ouvrons le menu *Text/ApplyLexical Resources* (sans oublier de cliquer sur le bouton *Clear*). La *Word List* recense 115 mots reconnus²⁰. Fermons la *Word List*.

²⁰ Ou plutôt, cent onze *simple word lexical entries* et quatre *compound lexical entries*.

3 Quelques remarques supplémentaires

3.1 Ajouter des traits

Les dictionnaires Unitex peuvent contenir des traits informatifs que le manuel nomme *codes sémantiques*. Ils suivent la catégorie et sont précédés par le signe + ; si la morphologie est présente, elle est située après les traits. Les traits ajoutés sur un dictionnaire de formes vedettes sont retranscrits dans le dictionnaire de formes fléchies.

Reprenons notre premier exemple légèrement modifié :

```
ordinateur,N100+object
fauteuil,N100
couloir,N100
```

Le dictionnaire de formes fléchies aura deux entrées avec un trait :

```
ordinateur,.N+object:ms
ordinateurs,ordinateur.N+object:mp
...
```

3.2 Créer directement des entrées fléchies

Il est possible d'ajouter des entrées directement dans le dictionnaire *MonDicoflx.dic* ou même de créer directement un fichier de formes fléchies. Ceci est particulièrement utile pour un mot qui ne se fléchit pas. Il est aussi possible²¹ de placer des commentaires en fin de ligne, après un /.

On pourrait par exemple compléter le dictionnaire *MonDicoflx.dic* par :

```
achète,acheter.V:P1s
...
voisins,voisin.N:mp
Phileas,.N+forename:ms
Fogg,.N+surname:ms/Héros du roman de Jules Vernes
```

3.3 Vérifier automatiquement le format du dictionnaire

Après des ajouts, nous pouvons vérifier que le dictionnaire obtenu est bien formé (c'est-à-dire, suit la forme attendue par Unitex). Cliquons sur le menu *DELA/Check Format* et sur le bouton *Check Dictionary*. Nous obtenons, dans le dossier *DELA*, un fichier nommé *CHECK_DIC.TXT* qui s'ouvre dans une fenêtre à l'écran.

Pour notre dictionnaire final, cela donne :

1. Des informations quantitatives : *344 lines read, 277 simple entries for 68 distinct lemmas, 67 compound entries for 15 distinct lemmas.*
2. La liste des caractères utilisés : *All chars used in forms... B (0042)... ô (00F4).*
3. Les codes grammaticaux ou sémantiques utilisés : *3 grammatical/semantic codes used in dictionary V N A.*

²¹ Malheureusement, ce n'est pas possible dans un dictionnaire de formes vedettes.

4. Les codes morphologiques : *10 inflectional codes used in dictionary P1s P3s... fs fp.*

3.4 Décrire un dictionnaire

Enfin, nous pouvons ajouter une description de notre dictionnaire dans un fichier *MonDicoflx.txt*. Par exemple : *Dictionnaire créé dans le cadre du tutoriel Unitex Dicos*. Cette information peut être lue dans le menu *Text/ApplyLexical Resources* par un clic-droit sur le nom du dictionnaire.