Manipulations avec les grammaires « Frantext »

Une partie de la base textuelle Frantext a été « catégorisée » : chaque mot est accompagné de sa catégorie syntaxique (verbe, nom, adjectif, etc.). Cela rend possible des recherches plus sophistiquées que dans la version brute.

En effet, plutôt que de spécifier la séquence recherchée uniquement par les mots, on peut faire directement référence à une classe de mots d'une catégorie donnée. Par exemple, la séquence `un &e(g=A) homme` permettra de retrouver tous les contextes où apparaît une expression comme *un grand homme*, *un bel homme*, *un petit homme*, etc.

On pourra même utiliser un dispositif plus élaboré pour affiner les recherches : une « grammaire ». Voici, à titre d'exemple, une grammaire que nous appellerons `g1`.

```
quantifieur :
   très | assez | excessivement
qualifieur :
   grand | gros | petit
qualification :
   &rqualifieur | &rquantifieur &rqualifieur
```

Par le symbole `&r`, on peut faire appel à une des règles de la grammaire. Par exemple, si on recherche la séquence `&rquantifieur,g1`, on fait appel à la règle appelée `quantifieur` dans la grammaire `g1`, ce qui correspond donc à l'un des trois mots *très*, *assez*, *excessivement*.

On peut bien sûr utiliser dans les grammaires toutes les expressions déjà définies. Par exemple, on peut définir comme règle de grammaire (`Ps` = participe passé) :

```
temps_compose:
   (&cêtre | &cavoir) &e(g=Ps)
```

Manipulations

Le site Frantext catégorisé se trouve à l'adresse : `http://jupiter.inalf.cnrs.fr/categ.htm`. On trouvera de l'aide dans le menu « Création de grammaires ».

1. Faire une recherche des occurrences des verbes à l'infinitif niés (*ne pas manger*, *ne jamais se tromper*, *ne plus rien savoir*...). Créer pour cela une grammaire simple, que vous pourrez affiner.
2. Faire une recherche des groupes prépositionnels compléments de nom

Le **cycle de travail** est le suivant :

1. Sélection du corpus de travail
2. (Création/modification de grammaires)
3. Recherche dans le corpus de travail

La liste des **expressions de séquence** est la suivante :

une graphie donnée	`une maison blanche`
une expression de choix	`maison ( blanche | bleue )`
une liste (préalablement définie)	`&lcouleur`
un verbe conjugué	`&caimer`
un mot fléchi	`&cvert`
une sous-expression optionnelle	`un & ?très grand homme`
un mot quelconque	`&q`
un mot catégorisé	`&e(g=V)`
une règle de grammaire	`&rg1`

Liste des codes grammaticaux :

A	adjectif (sauf cas Aca, Apr, Aps)
Aca	adjectif cardinal
APr	adjectif/participe présent
APs	adjectif/part. passé
Adv	Adverbe
Cc	Conjonction coordination
Cs	Conjonction subordination
D	Déterminant (sauf cas Dca, Dg)
Dca	car. dét (cardinal ayant le rôle d'un déterminant : *deux pigeons s'aimaient*)
Dg	amalgamés (au/aux/du/des)
E	exclamatif
Ep	présentatif (voici, voilà, ...)
Ger	gérondif (en lié à un participe présent)
Inf	infinitif
Inj	interjection (ah, oh, ha, ho, ...)
Int	interrogatif
Np	Nom propre
Nu	numeral card.
Ono	onomatopée
P	Pronom (sauf cas Per, X)
Per	Pronom personnel
Pp	Préposition
Pr	Participe présent sauf cas APr, Ger
Ps	Participe passé (sauf cas APs)
S	Substantif
V	Verbe (sauf participes et infinitif)
R	mot inconnu du logiciel
X	mot non traité (que/qu', où, sinon)