

Introduction au TAI

Le TAI est un processus linguistique qui permet de créer de nouveaux mots à partir d'un mot de base.

Il existe deux types de TAI : le TAI morphologique et le TAI dérivationnel.

Morphologie

La morphologie est l'étude des unités de sens (les morphèmes) et de leur combinaison pour former des mots.

Les morphèmes peuvent être classés en deux catégories : les morphèmes libres et les morphèmes liés.

Les morphèmes libres sont des mots qui peuvent apparaître seuls dans une phrase. Les morphèmes liés sont des unités qui ne peuvent apparaître que dans un mot.

Les morphèmes liés sont classés en deux catégories : les affixes et les radicaux.

Les affixes sont des unités qui sont ajoutées à un mot de base pour en modifier le sens ou la fonction grammaticale.

Les radicaux sont des unités qui sont ajoutées à un mot de base pour en modifier le sens.

Les affixes peuvent être classés en deux catégories : les affixes grammaticaux et les affixes lexicaux.

Les affixes grammaticaux sont des unités qui sont ajoutées à un mot de base pour en modifier la fonction grammaticale.

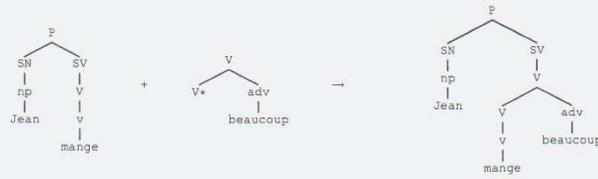
► **Exercice 2. Lexiques et morphologie dérivationnelle** Le développement manuel de ressources lexicales est une tâche fastidieuse. Toute approche permettant d'automatiser tout ou partie du processus est donc la bienvenue — mais nécessite en général une validation manuelle.

1. Rappeler brièvement ce qu'est un lexique morphologique, un lexique syntaxique et un lexique sémantique.
2. La **dérivation morphologique** est un processus linguistique permettant de créer un mot à partir d'un mot de base et de **affixes** (en français, de préfixes ou de suffixes). Proposez une analyse en base et affixes pour les mots suivants. On indiquera dans chaque cas la catégorie du mot de base et celle du mot dérivé.
rejouer, surentraînement, déconstruction, décomplexer, désamiantage, chauffagiste, droit-de-l'hommisme, insoluble, réévaluation
3. Chacun des affixes listés à la question précédente a-t-il un sens similaire d'un exemple à l'autre ? Indiquez quel est cet effet sémantique. Par exemple, le suffixe *-ifier* peut vouloir dire, entre autres, *rendre ADJ (complexe → complexifier = rendre (plus) complexe)* ou *transformer en NC (vin → vinifier = transformer en vin)*.

1. C. B. F. G. H. I. J. K. L. M. N. O. P. Q. R. S. T. U. V. W. X. Y. Z.

1. C. B. F. G. H. I. J. K. L. M. N. O. P. Q. R. S. T. U. V. W. X. Y. Z.

Adjonction :



On appelle *ancree* une feuille qui n'est ni un nœud pied ni un nœud de substitution : c'est alors un nœud représentant un mot.

On rappelle enfin que l'ensemble des opérations permettant de construire l'arbre complet d'une phrase à partir des arbres élémentaires peut être représenté par un arbre dit *arbre de dérivation* (l'arbre de la phrase s'appelle *arbre dérivé*). Ainsi, l'arbre de dérivation représentant l'analyse de la phrase *Jean mange beaucoup* au moyen de la séquence des deux opérations ci-dessus est représentée par l'arbre de dérivation suivant :



Cet arbre indique que l'arbre d'ancree Jean est substitué dans celui d'ancree mange, et que l'arbre d'ancree beaucoup y est adjoint. On fait en général en sorte, autant que possible, que l'arbre de dérivation corresponde à un arbre sémantique (le nœud correspondant à un prédicat a pour fils ses arguments et ses modifieurs/circonstants).

seront les feuilles des arbres de la grammaire. Par ailleurs, nous allons construire une grammaire telle que chaque arbre élémentaire contienne un et un seul nœud ancree.

On suppose que si un nœud N a la même étiquette que son père, c'est qu'il y a eu adjonction d'un arbre auxiliaire. Identifier dans les arbres du mini-corpus les endroits où l'on trouve cette configuration. Combien d'arbres auxiliaires sont-ils en jeu ? Dessinez cet/ces arbre(s) auxiliaire(s), ainsi que l'arbre des phrases correspondantes avant que la/les adjonction(s) ai(en)t eu lieu.

3. Une fois les arbres auxiliaires identifiés et retirés des arbres du mini-corpus, nous allons « découper » ces arbres en arbres initiaux. Proposer un inventaire d'arbres initiaux couvrant exactement le mini-corpus arboré, en indiquant le nombre de fois que chacun d'eux est utilisé dans le mini-corpus. On prendra garde à faire en sorte que lorsque l'on reconstruit l'arbre complet d'une phrase, l'arbre de dérivation obtenu soit satisfaisant (le nœud pour l'arbre ancree le verbe principal doit par exemple être la racine de l'arbre de dérivation).
4. Normalisez les comptes ainsi obtenus sur les arbres initiaux pour en faire des probabilités de réécriture : la somme des probabilités des arbres initiaux de même racine doit être égale à 1.
5. Que proposez-vous comme probabilités pour le(s) arbre(s) auxiliaire(s) ?
6. Calculez les probabilités globales de chacun des quatre arbres du mini-corpus.
7. Soit la phrase *Pierre commence à dîner*. Complétez le lexique pour qu'il couvre tous les mots de cette phrase (on n'oubliera pas que *dîner* est à la fois un nom commun et un infinitif).
8. Combien d'analyses cette phrase a-t-elle selon la grammaire d'arbres adjoints que vous avez extraite précédemment du mini-corpus arboré ? Quelles sont les probabilités associées à ces analyses ? L'analyse la plus probable est-elle la bonne ?
9. Quelle est votre impression générale sur cette approche ? Quelles idées auriez-vous pour la rendre plus sophistiquée ? Quels avantages et/ou inconvénients lui trouvez-vous par rapport aux PCFG ?