

Annotation sémantique des morphèmes

Introduction

L'objectif de cette session est de tester l'applicabilité des catégories sémantiques définies pour l'annotation des noms (en temps que mot) à une échelle plus petite, celle des morphèmes nominaux, tels que les suffixes et préfixes en français.

Nous avons préparé un corpus de 100 noms dérivés, chacun accompagné de sa définition et de son procédé morphologique. Votre tâche consiste à annoter les morphèmes nominaux en utilisant les catégories sémantiques pré-définies (Huguin et al., 2023) dans le cadre d'un autre projet (Namer et al., 2023). Pour ce faire, vous devrez examiner chaque morphème, comprendre la définition du mot auquel il appartient, et identifier la catégorie sémantique appropriée à partir du guide d'annotation fourni.

L'annotation se fera en temps réel, c'est-à-dire que vous devrez effectuer cette analyse et sélection au fur et à mesure, sans préparation préalable pour chaque mot. Ce processus est conçu pour durer une matinée au maximum, permettant ainsi une évaluation rapide et efficace de la faisabilité de notre approche sur ces unités linguistiques plus petites.

Le guide d'annotation sémantique

Le guide d'annotation que vous allez utiliser est structuré selon une hiérarchie claire de catégories sémantiques et leurs sous-catégories. En introduction, vous trouverez un aperçu de cette hiérarchie qui vous aidera à naviguer parmi les différentes options.

En plus des catégories simples, le guide comprend également des catégories complexes, qui combinent deux catégories simples de deux manières distinctes :

- **Avec l'opérateur "x"** : Ces catégories sont utilisées pour annoter des noms nécessitant les étiquettes "group" et "part" associées à une autre étiquette, afin de capturer des nuances sémantiques spécifiques.

Si le nom est monosémique, il se voit attribuer une étiquette simple ou une étiquette complexe manipulant l'opérateur x. Par exemple *bavard* est annoté Person. Les étiquettes Group ('un ensemble de') et Part ('une partie de') peuvent être associées à une autre étiquette simple avec l'opérateur x qui exprime la distribution (A de B). Ainsi, le lexème *fôret* est annoté GroupxPlant.

- **Avec l'opérateur "+"** : Ces catégories permettent d'annoter des noms à plusieurs sens compatibles, c'est-à-dire des noms qui possèdent plusieurs facettes sémantiques en un.

Si le nom possède plusieurs sens compatibles appelés hybrides (vs. les sens sont incompatibles ou antagonistes en discours appelés noms polysémiques). Les noms à sens multiples compatibles, dits hybrides, admettent la co-prédication, cf. (1). [...] [L']opérateur + qui exprime la conjonction (A et B). En (1a), le locuteur se focalise sur le contenu du livre (A), en (1b) sur sa matérialité (B), et (1c) concilie les deux lectures (A et B) [...]. Le lexème nominal **livre** est donc annoté **Artifact+Cognition**.

- (1) a. Ce **livre** est passionnant. **[Cognition]**
- b. Ma mère m'a acheté un **livre** magnifiquement relié. **[Artifact]**
- c. J'aime ce **livre**. **[Artifact+Cognition]**

Chaque catégorie et sous-catégorie est décrite en détail, avec des définitions précises et, souvent, des tests linguistiques pour vous guider. Ces tests prennent la forme de phrases à trou où le nom avec le morphème annoté peut être inséré, vérifiant ainsi la pertinence de la catégorie choisie.

Il est à noter que certains tests sont **conjonctifs** (et), c'est-à-dire que pour être considérés comme positifs, le nom testé doit entrer dans l'ensemble des structures distributionnelles, et que d'autres sont **disjonctifs** (ou).

Corpus et méthode d'annotation

Le corpus que vous allez annoter se compose de 100 noms dérivés. Il est disponible aux formats .xlsx et .ods. Chaque entrée du corpus inclut :

- **Le morphème (suffixe ou préfixe)** : Le morphème à annoter.
- **Le mot dérivé** : Le nom dans lequel on trouve le morphème en contexte.
- **Le procédé morphologique** : Une description simple du processus de dérivation utilisé pour former le mot.
- **La définition** : Une explication succincte du sens du mot.

Morphème	Nom dérivé	Procédé dérivationnel	Définition	Étiquettes	Commentaires
-ier	bijoutier	bijou + (t) + ier	Personne qui vend des bijoux		

Ensuite, vous trouverez les cellules vides à remplir. Dans un premier temps, la ou les catégorie(s) sémantique(s) que vous avez identifiée(s), puis une cellule pour les commentaires.

Voici les possibilités d'annotation auxquelles vous ferez, *a priori*, face :

- Réponse sûre : une catégorie correspondante
- Réponse incertaine : plusieurs catégories correspondantes
- Pas de réponse : aucune catégorie correspondante

Les commentaires permettent de compléter votre réponse et exprimer votre ressenti sur l'annotation.

Comme mentionné précédemment, les catégories sémantiques sont organisées selon une hiérarchie. Lorsque vous choisissez une catégorie, il est crucial de sélectionner celle qui est la plus fine et donc située au plus bas de la hiérarchie, à condition qu'elle corresponde bien au morphème en question.

⚠ Nous vous remercions par avance pour votre rigueur et votre attention. Vos annotations seront cruciales pour vérifier si les catégories sémantiques actuelles peuvent être étendues efficacement aux morphèmes nominaux, et pour ajuster nos méthodologies en conséquence.

Une fois l'annotation terminée, vous pourrez enregistrer le corpus annoté à **votre nom** et l'envoyer à l'adresse suivante corentin.longueepée@edu.univ-fcomte.fr avec en Cc izabella.thomas@univ-fcomte.fr et yagmur.ozturk@edu.univ-fcomte.fr.

Bon travail à tous et n'hésitez pas à poser des questions si vous avez besoin de clarifications pendant la session.