

# ETUDE COMPUTATIONNELLE DE LA STABILITÉ INTERLANGUE DES CATÉGORIES MORPHO-SYNTAXIQUES

Rapport de Stage de L3

Matthieu Boyer

21 juin 2024



## Table des matières

1	Why ?	1
2	Première Approche	1

### Résumé

Dans ce rapport nous nous intéressons à la stabilité interlangue des catégories morpho-syntaxiques. Nous avons quantifié la manière dont différentes catégories descriptives d'un langage ont différentes significations dans différents langages, et particulièrement la manière dont un concept est matérialisé dans différents langages.

## 1 Why ?

Martin Haspelmath sur la différence entre une catégorie linguistique descriptive dans un langage et une catégorie linguistique comparative dans le méta-langage :

There is a fundamental distinction between language-particular categories of languages (which descriptive linguists must describe by descriptive categories of their descriptions) and comparative concepts (which comparative linguists may use to compare languages).

*Martin Haspelmath, HOW COMPARATIVE CONCEPTS AND DESCRIPTIVE LINGUISTIC CATEGORIES ARE DIFFERENT* Dans ce rapport, nous allons donc nous intéresser à la notion fondamentale de catégorie morphosyntaxique, et comparer les descriptions dans différents langages de catégories linguistiques comparatives. Pour ce faire, nous allons considérer que les relations de dépendances (*reldep*) décrites par les annotations de UNIVERSAL DEPENDENCIES (UD) sont une manière de représenter des catégories comparatives.

## 2 Première Approche

Nous considérons tout d'abord que chaque *reldep* décrit une unique catégorie comparative et que plusieurs *reldep* ne peuvent instancier une même catégorie comparative. En comptant le nombre

d'instances de chaque *reldep* pour un mot vérifiant une propriété grammaticale de la langue (donc une catégorie descriptive, que l'on représente par une *feature* de UD, typiquement les cas pour des langues en utilisant), on obtient une représentation vectorielle des catégories descriptives et on peut donc mesurer la proximité de deux catégories descriptives dans deux langues différentes en utilisant par exemple la distance cosinus. On trouve par exemple les résultats suivants :

Proximity with :	Case=Voc	Case=Nom	Case=Dat	Case=Acc	Case=Gen	Case=Loc
Median	0.0	0.0	0.0	0.0	0.0	0.0
Mean	0.05971	0.14069	0.10222	0.19691	0.10808	0.08202
Low	88304	202584	141216	117296	179856	91820
High	38517	82780	46133	125164	51907	45144

TABLE 1 – Proximities for Case=Acc

Proximity with :	Case=Voc	Case=Nom	Case=Dat	Case=Acc	Case=Gen	Case=Loc
Median	0.0	0.0	0.0	0.0	0.0	0.0
Mean	0.03287	0.07652	0.09264	0.07748	0.06972	0.06423
Low	72292	152242	112560	116542	118074	68894
High	16906	40131	38632	32442	27837	33811

TABLE 2 – Proximities for Case=Dat

Proximity with :	Case=Voc	Case=Nom	Case=Dat	Case=Acc	Case=Gen	Case=Loc
Median	0.0	0.0	0.0	0.0	0.0	0.0
Mean	0.04438	0.10685	0.0824	0.10114	0.13129	0.06767
Low	93432	189382	139810	163898	164464	98714
High	29999	62897	35412	49688	88856	36335

TABLE 3 – Proximities for Case=Gen

Proximity with :	Case=Voc	Case=Nom	Case=Dat	Case=Acc	Case=Gen	Case=Loc
Median	0.0	0.0	0.0	0.0	0.0	0.0
Mean	0.01854	0.03652	0.04418	0.03796	0.03914	0.0558
Low	29488	62408	43138	48224	49754	39600
High	13877	23226	24074	23148	20250	42232

TABLE 4 – Proximities for Case=Loc

Proximity with :	Case=Voc	Case=Nom	Case=Dat	Case=Acc	Case=Gen	Case=Loc
Median	0.0	0.0	0.0	0.0	0.0	0.0
Mean	0.07986	0.26295	0.10767	0.15292	0.12731	0.09068
Low	90374	73616	195126	205522	213362	120614
High	62653	216768	64069	89263	79326	63867

TABLE 5 – Proximities for Case=Nom

Proximity with :	Case=Voc	Case=Nom	Case=Dat	Case=Acc	Case=Gen	Case=Loc
Median	0.0	0.0	0.0	0.0	0.0	0.0
Mean	0.03415	0.01876	0.01431	0.01622	0.01667	0.01231
Low	5608	28136	28046	26906	24358	17678
High	33196	15544	10214	12265	12275	10258

TABLE 6 – Proximities for Case=Voc