

La constitution d'un corpus d'ellipses erronées pour la représentation et la transmission des connaissances

Laura NORESKAL

Doctorante en linguistique – traitement automatique du langage

Laboratoire MODYCO – UMR 7114 & Université Paris Nanterre

laura.noreskal@parisnanterre.fr



- Projet national impliquant 13 universités françaises
- Utiliser le numérique pour développer de nouvelles formes d'évaluation et de formation
- Développement d'un dispositif national d'évaluation, de formation et de certification des compétences d'expression et de compréhension écrites en français
- Plusieurs modules médiatisées déjà déployés notamment à Paris Nanterre

Les constructions elliptiques

- Premier repérage des zones de difficultés rédactionnelles : Les constructions elliptiques sont souvent sujettes aux erreurs dans les rédactions des étudiants
- Plusieurs modules pour déterminer les contextes syntaxiques favorisant l'ellipse et plus exactement les erreurs
- Étude linguistique pour la détection automatique

La détection automatique des ellipses

Objectifs :

- Détecter automatiquement les ellipses dans des textes brefs d'étudiants
- Reconnaître les erreurs de constructions
- Proposer un feed-back

Les ellipses

- Dans un énoncé constitué de deux propositions A et B liées par une **relation sémantique**, une ellipse est l'**omission** dans B d'**éléments récupérables** grâce à A qui fournit un **contexte direct**. La proposition B est **compréhensible** grâce aux informations données par les **éléments réalisés** dans A.
- Selon Abeillé & Mouret (2008) : « On parle d'ellipse [dans les coordinations elliptiques] lorsque l'interprétation d'une forme syntaxique **requiert plus que ce qui est fourni** par les éléments qui la composent et que le **matériel nécessaire pour obtenir cette interprétation est récupérable** dans le contexte immédiat. »
- Plusieurs types d'ellipses ont été étudiés tels que :
 - Le gapping (Ross 1970) : Jean boit du thé et Marie du café
 - Les ellipses polaires : Marie viendra et Pierre aussi

Les ellipses (2)

- Les recherches ont montré que les ellipses répondent à des contraintes linguistiques qui assurent la stabilité d'un énoncé. Le non-respect de ces contraintes entraîne une instabilité et donc une incompréhension de l'énoncé. (Abeillé 2008, Desmets 2008, Bîlbîe 2011)
 - « L'ellipse est un phénomène d'interface qui implique des contraintes de parallélisme syntaxique, sémantique et pragmatique » (Desmets, 2008)
- (1) *Ces références dépendent du contexte de la phrase mais aussi la sémantique.
- (2) Ces références dépendent du contexte de la phrase mais aussi de la sémantique.

La constitution d'un corpus d'erreurs d'ellipses

Objectifs :

- Repérer le type d'erreurs d'ellipses le plus fréquent dans les rédactions d'étudiants
- Déterminer le type de rédactions le plus sujet aux erreurs
- Constituer un corpus d'entraînement pour notre outil de détection automatique

Méthodologie



Collecte de plusieurs rédactions d'étudiants



Catégorisation des rédactions



Extraction des erreurs d'ellipse



Choix des métadonnées à garder

Collecte des rédactions d'étudiants

- Examens
- Exercices
- Devoirs maison
- Rapports de stage
- Mémoires



Catégorisation des rédactions

Rédactions spontanées :

- Examens
 - Exercices
- Temps limité, peu de relecture

Rédactions préparées :

- Devoirs maison
 - Rapports de stage
 - Mémoires
- Temps de préparation plus long et possibilité de relecture par l'étudiant ou par un tiers (proche, professeur, directeur de recherches...)



Extraction des erreurs d'ellipse

- Lecture des rédactions dans leur totalité
- Extraction manuelle des erreurs avec un contexte gauche de 2 à 3 phrases et un contexte droit d'une phrase



Les métadonnées

Pour l'étude du phénomène, nous avons gardé :

- La nature de la rédaction
- Le niveau d'étude
- Le domaine d'étude
- La rédaction dans sa totalité



Corpus

Extraits	Type de rédaction	Niveau	Domaine
[...] Il s'agit soit d'un couple hétérosexuel stérile ou ayant des problèmes de fertilité, soit <err> un couple homosexuel ne pouvant, par nature, pas avoir d'enfant.	Exercice	L1	LLCER Anglais
[...] Le client a été blessé et n'a pas eu de soins. Il a reçu des coups de matraques, l'interdiction de faire ces besoins et <err> privé de nourriture.	Rapport de stage	M2	Droit
[...] La référence est générique car on parle de toutes les salades et pas <err> une en particulier.	Examen	L1	Sciences du langage
[...] Toutes ces dispositions permettent à l'élève de réussir et donc <err> essayer d'augmenter l'égalité des chances.	Devoir maison	M1	MEEF

Premières observations

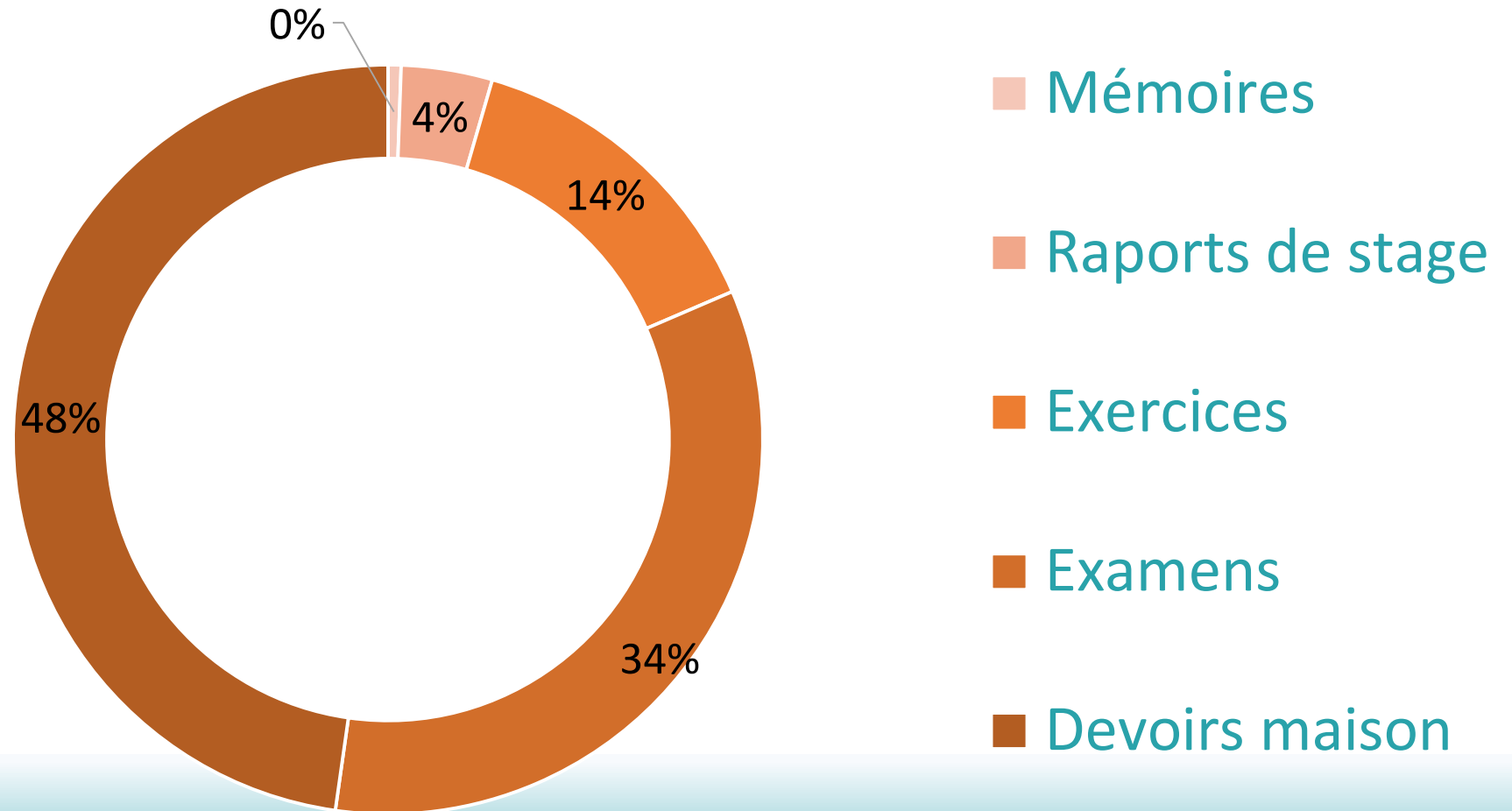
Nous avons collecté 250 productions dont **178** erreurs d'ellipse :

Type de rédactions	Nombre de redactions	Taux d'erreurs
Rédactions spontanées	160	85 (47,8%)
Rédactions préparées	90	93 (52,2 %)
Total	250	178

Premières observations (2)

- **Rédactions spontanées:** 160 productions dont **85** erreurs (**47,8 %**)
 - Examens :** 119 productions dont **60** erreurs
 - Exercices :** 41 productions dont **25** erreurs
- **Rédactions préparées :** 90 productions dont **93** erreurs (**52,2 %**)
 - Devoirs maison :** 82 productions dont **85** erreurs
 - Rapports de stage :** 7 productions dont **7** erreurs
 - Mémoire :** 1 production dont **1** erreur

Premières observations (3)



Perspectives

- ✓ Constituer un corpus de 250 à 300 erreurs d'ellipses et d'ellipses bien formées
- ✓ Mise en place d'un test sur 250 étudiants « provoquant » la création d'ellipses
- ✓ Analyser le phénomène en prenant en compte le contexte (les verbes impliqués, la place de l'erreurs dans la rédaction...)
- ✓ Tester les différentes méthodes du traitement automatique du langage

Merci !

References

- Abeillé A., Mouret F. : Quelques contraintes sur les coordinations elliptiques en français. *Revue de sémantique et de pragmatique* (2008)
- Bilbiie G. : Grammaire des constructions elliptiques. Une étude comparative des phrases sans verbe en roumain et en français, Thèse de Doctorat, Université Paris 7 (2011)
- Desmets M. : L'ellipse dans les constructions en comme, *LINX* (2008)
- Mouret F. : Grammaire des constructions coordonnées en français, Thèse de Doctorat, Université Paris 7 (2007)
- Ross J. R.: Gapping and the order of constituents. In Manfred Bierwisch & Karl Erich Heidolph (eds.), *Progress in linguistics* (1970) 249-259