

## Corpus écrits et transcrits

### TP Inception

#### NORM-EC-CE2-2016-98-D1-S940-V1

Le problème ici est de déterminer qui est le sujet du verbe dans « *vit qu'elle était magique, elle parlait !* ». Le verbe « voir », conjugué à la 3<sup>ème</sup> personne du passé simple, est utilisé dans une phrase où le sujet n'est pas explicité, correspondant à une anaphore zéro. La question qui se pose est de savoir si le sujet de l'action « elle en sortit » reste le même (c'est-à-dire la chatte avec l'étiquette « cat »), suggérant ainsi qu'elle découvre ses pouvoirs. Il est cependant possible que le verbe « vit » réfère au loup, ce qui sous-entend qu'il aperçoit les pouvoirs magiques de la chatte. Nous avons toutes les deux choisi de coréférencer « vit » à « cat ».

Lors de l'adjudication, une difficulté est apparue concernant l'expression « son ami le chasseur » : doit-on considérer le groupe comme une seule unité, ou bien séparer « son ami » et « le chasseur » ? Après discussion, il est apparu plus logique de les traiter séparément, car le déterminant défini « le » fait la distinction entre les deux.

Notre score inter-annotateur est de 0,86, ce qui est considéré comme assez satisfaisant.

#### NORM-EC-CE2-2016-98-D1-S936-V1

La détermination du sujet tout au long du corpus a été difficile en raison d'un emploi problématique du discours direct. Les dialogues s'enchaînent, et en cherchant à maintenir une certaine cohérence, nous constatons que les guillemets utilisés pour introduire le discours direct ne sont pas correctement appliqués. Deux chats sont présents dans le texte, mais il est difficile de distinguer qui est qui. Nous comprenons que l'un des chats (cat) menace l'autre (cat1) avant de révéler qu'il (cat) plaisante. Son interlocuteur (cat1) s'apprête à partir et est interpellé, ne saisissant pas l'expression « vieil ami ». L'autre chat (cat) lui répond qu'il n'a pas compris. Le verbe « voir », conjugué à la 3<sup>ème</sup> personne du passé simple, est utilisé dans une phrase coordonnée où le sujet n'est pas explicité, correspondant à une anaphore zéro. Des verbes introducteurs de parole devant, après ou à l'intérieur des paroles rapportées auraient permis une meilleure compréhension. Nous avons eu des difficultés sur le fait que le pronom « tu » de la ligne 4 fait référence au second chat, établissant une coréférence assez déroutante avec le « je ». La phrase suivante est construite : « (cat) : attends, mon vieil ami, (cat1) : je ne suis pas vieux, (cat) : tu n'as pas compris le mot ».

Lors de l'adjudication, nous avons été surprises de ne pas avoir de désaccord, bien que ce texte ait été l'un de ceux qui nous ont le plus compliqué la tâche lors de l'annotation.

Notre score inter-annotateur est de 1, ce qui est considéré comme parfait.

#### NORM-EC-CE2-2016-98-D1-S834-V1

Il n'y a pas eu de difficulté particulière pour l'annotation de ce corpus. Les deux référents, la chatte et le robot, sont identifiés clairement. Concernant la coréférence, les

pronoms « il » et « ils » n'ont pas posé de problème, l'un renvoyant au robot, et l'autre à la chatte et au robot.

Lors de la phase d'adjudication, nous avons eu des différences concernant l'annotation de l'infinitif passé « avoir réfléchi », à savoir s'il est bien dans une phrase où le sujet n'est pas explicité avant. Nous avons conclu que c'était le cas et qu'il est coréférencé au robot, car ensuite, le même sujet est suivi d'un autre verbe « dit ». De même pour « l'amena », où le verbe est aussi coréférencé au robot.

Notre score inter-annotateur est de 0,95, considéré comme assez satisfaisant.

### **NORM-EC-CE2-2016-98-D1-S816-V1**

« *Et il courut en direction des chasseurs. Le loup, qui attendait sa réponse, finit par le pourchasser.* » Le pronom personnel « il » entraîne une ambiguïté. Il est difficile de savoir dans l'immédiat si c'est le robot ou le loup qui court en premier. Ce n'est qu'en se référant à la phrase suivante que cela devient plus clair. La présence de nombreux personnages avec les animaux, le robot, le loup et les chasseurs, complique le suivi des référents. Six étiquettes différentes pour les entités animées non indiquées par la consigne ont été utilisées. La première mention de la biche accompagnée du cerf et du renard semble correspondre aux mêmes animaux que le loup veut manger plus tard. Mais le texte n'étant pas explicite à ce sujet, on pourrait aussi interpréter que le loup parle de ces animaux de manière générale, sans se référer spécifiquement à ceux rencontrés par le robot. Cela dépend alors de l'interprétation, mais en l'absence de détails supplémentaires il s'agit probablement des mêmes animaux.

Lors de la phase d'adjudication, la métaphore « un tas de ferraille » a suscité un débat, à savoir si elle était coréférencée avec le robot, ne désignant pas une entité animée. Nous avons finalement interprété cette expression comme un surnom péjoratif du robot lui-même. Le groupe nominal « toutes sortes d'animaux » a aussi été discuté, cette fois en tant que cataphore coréférencée aux animaux « cerf, biche, coccinelle, renard ». L'annotation du verbe « finit » coréférencée avec le loup, facilite la compréhension de la phrase.

Notre score inter-annotateur est de 0,35, considéré comme mauvais.

### **NORM-EC-CE2-2016-96-D1-S1926-V1**

« *La sorcière disait : « Non, vous vous trompez de chat », et la sorcière claquait la porte.* » Il y a une ambiguïté autour du 1<sup>er</sup> « vous », le pronom personnel sujet à la deuxième personne du pluriel. La sorcière pourrait s'adresser à la fois au père et à la fille ou seulement à l'un d'eux. Il semble plus cohérent que l'emploi du « vous » indique plusieurs personnes, en faisant référence aux derniers personnages mentionnés. En privilégiant la fille, nous tenons compte du fait qu'elle est la dernière à parler. Mais il est aussi possible que la sorcière utilise le « vous » pour s'adresser à un adulte en général, laissant la place au père. Le choix du « vous » avec le père et la fille semble plus cohérent. Ensuite, le policier mentionné dans le texte est coréférencé avec les policiers. Il s'agit d'un membre du groupe de policiers qui ont frappé à la porte. « *La petite fille criait : « Nougat » parce que le chat s'appelait comme ça.* » Ici, l'expression « comme ça » n'est pas coréférencée au chat, car elle rappelle simplement le nom que porte le chat, « Nougat ».

Lors de la phase d'adjudication, le policier mentionné a suscité un débat dans le texte, à savoir s'il est coréférencé avec les policiers. Il (ext4) s'agit d'un membre du groupe de policiers

(ext3) qui ont frappé à la porte. Nous avons également convenu d'annoter « Nougat » en le liant à son adjectif « seul », formant ainsi l'expression « Nougat seul ».

Notre score inter-annotateur est de 0.93, considéré comme assez satisfaisant.

#### **NORM-EC-CE2-2016-96-D1-S1866-V1**

Ici, nous n'annotons pas l'énumération « des souris, des crapauds », car elle ne fait pas référence à des souris ou des crapauds en particulier qui apparaissent dans le texte, mais aux souris et crapauds en général. L'expression « petit rat » peut être interprétée de deux façons : soit comme une insulte, soit comme une confusion de la sorcière qui prend le chat pour un rat. Le chat, quant à lui, ne comprendrait pas cette méprise et lui répondrait en affirmant qu'il est bien un chat. Dans tous les cas, « petit rat » est coréférencé à (cat). Nous n'annotons pas « un rat » et « un chat », car cela fait référence à l'animal en général.

Lors de l'adjudication, nous avons eu une différence à la fin du texte. Il y a un référent évolutif : on annote la deuxième expression « en crapaud » en utilisant la même étiquette que celle utilisée pour le maillon de l'entité transformée. Mais après concertation, il n'y a pas d'annotation avec le premier « en crapaud » (ligne 11) car (cat) n'est pas encore transformé.

Notre score inter-annotateur est de 0.94, considéré comme assez satisfaisant.

#### **NORM-EC-CE2-2016-91-D1-S688-V1**

Pour les pronoms personnels clitiques « *le* » et « *lui* » à la ligne 2, ce n'est pas clairement explicité à qui ils font référence, que ce soit au chat ou au loup. Nous supposons que l'enfant a continué sa phrase en utilisant le chat comme sujet principal et nous interprétons « *le* » comme se référant au loup, et « *lui* » désignant le chat. L'utilisation excessive des pronoms personnels « *je* » et « *tu* » avec des pronoms personnels clitiques « *te* » et « *le* » complique l'annotation. Même si le dialogue direct facilite l'interaction entre les personnages, il rend l'identification des référents difficile. Le « *tu* » de la ligne 9 n'est pas immédiatement identifiable. Ce n'est qu'à la fin du texte que l'on comprend qu'il s'agit du chat qui griffe le loup (et non l'inverse). Des verbes introducteurs de parole permettraient une meilleure compréhension.

Lors de la phase d'adjudication, il n'y a pas eu de soucis majeurs.

Notre score inter-annotateur est de 0,97, considéré comme assez satisfaisant.

#### **NORM-EC-CE2-2016-91-D1-S1164-V1**

Le pronom personnel sujet à la 3<sup>ème</sup> personne du singulier de la 5<sup>ème</sup> ligne, « *elle* », ne permet pas de déterminer si c'est la fille ou la sorcière qui se cache. C'est seulement dans la suite de la phrase que nous comprenons à qui il se réfère.

Lors de la phase d'adjudication, il n'y a pas eu de soucis majeurs.

Notre score inter-annotateur est de 0.99, considéré comme assez satisfaisant.

#### **NORM-EC-CE2-2016-91-D1-S1160-V1**

Le pronom personnel clitique « *la* dévora » dans le texte peut prêter à confusion. On ne sait pas s'il fait référence à la grand-mère ou à la petite fille. Sans connaître l'histoire du « Petit Chaperon Rouge » en amont, il est difficile de déterminer à qui le pronom se rapporte. Nous savons donc que ce « *la* » réfère à la grand-mère. Puis, quand le texte rapporte que « le loup l'avalait », cela reste ambigu. On ne sait pas si c'est la grand-mère qui est avalée après avoir été

dévorée, ou si c'est la petite fille qui subit le même sort. Cette incertitude complique l'annotation en coréférence, car le référent n'est pas précisé.

Lors de la phase d'adjudication, nous avons discuté de l'annotation du verbe à l'impératif « fais » avec « attention », car ils forment ensemble une expression figée. Elle signifie « sois prudente ».

Notre score inter-annotateur est de 0.98, considéré comme assez satisfaisant.

### NORM-EC-CE2-2016-91-D1-S1155-V1

Il n'y a pas eu de difficultés particulières lors de l'annotation.

Lors de l'adjudication, nous nous sommes demandé si le groupe nominal « mon amie » pouvait être coréférencé avec (witch). Après réflexion, cela semble peu probable, car la sorcière n'a pas encore été explicitement désignée comme telle à ce moment-là.

Notre score inter-annotateur est de 0.97, considéré comme assez satisfaisant.

### CONCLUSION

Les pronoms personnels sujets (« je », « tu », « il », « elle ») ou clitiques (« le », « la », « lui »), compliquent très souvent l'annotation, en particulier dans les passages en discours direct et quand les personnages sont du même genre. Le manque de diversité dans les pronoms crée une confusion, car nous n'avons pas de moyen clair pour attribuer chaque pronom à un personnage spécifique. L'ambiguïté rend l'annotation de coréférence complexe et peut engendrer des erreurs dans l'interprétation des personnages. Il est également important de prendre en compte toutes les anaphores zéro possibles, ce qui n'est pas toujours un réflexe.

Mean	Min	Max	Variance
0.8946	0.3502	1.0000	0.0381

Agreement		
Document	Score	Annotators
NORM-EC-CE2-2016-91-D1-S1155-V1.conllu	0.97	Marion Leteillier, Naja Goncalves Dos Santos
NORM-EC-CE2-2016-91-D1-S1160-V1.conllu	0.98	Marion Leteillier, Naja Goncalves Dos Santos
NORM-EC-CE2-2016-91-D1-S1164-V1.conllu	0.99	Marion Leteillier, Naja Goncalves Dos Santos
NORM-EC-CE2-2016-91-D1-S688-V1.conllu	0.97	Marion Leteillier, Naja Goncalves Dos Santos
NORM-EC-CE2-2016-96-D1-S1866-V1.conllu	0.94	Marion Leteillier, Naja Goncalves Dos Santos
NORM-EC-CE2-2016-96-D1-S1926-V1.conllu	0.93	Marion Leteillier, Naja Goncalves Dos Santos
NORM-EC-CE2-2016-98-D1-S816-V1.conllu	0.35	Marion Leteillier, Naja Goncalves Dos Santos
NORM-EC-CE2-2016-98-D1-S834-V1.conllu	0.95	Marion Leteillier, Naja Goncalves Dos Santos
NORM-EC-CE2-2016-98-D1-S936-V1.conllu	1.00	Marion Leteillier, Naja Goncalves Dos Santos
NORM-EC-CE2-2016-98-D1-S940-V1.conllu	0.86	Marion Leteillier, Naja Goncalves Dos Santos

Tout d'abord, nous avons analysé l'utilisation des pronoms clitiques objet direct et indirect « *me, te, se, nous, vous, le, la, lui, les, leur* » dans les productions écrites d'enfants. Nous supposons que l'emploi de ces pronoms entraînerait des ambiguïtés en compliquant la cohérence narrative. Sur les 20 pronoms analysés avec principalement « *me, te, le, la* », seules 4 occurrences étaient ambiguës (20%). Ces résultats montrent une maîtrise générale des pronoms clitiques, même s'il existe des difficultés dans des contextes ambigus.

Ensuite, nous avons étudié les anaphores zéro, où les enfants omettent un pronom sujet et souvent après des mots comme « *et* » ou « *après* ». Nous partions du principe qu'une utilisation fréquente de ces anaphores zéro compliquerait la compréhension du texte. Mais sur les 14 anaphores zéro identifiées, une seule était ambiguë. Ces omissions ne génèrent généralement pas de confusion dans les textes des enfants.

Enfin, nous avons examiné la relation entre la longueur des textes et la clarté des références. Nous avons supposé que des textes plus longs, avec plus de personnages, augmenteraient la complexité narrative et aussi les risques de confusion. Mais même s'il peut y avoir des ambiguïtés aussi bien dans les textes courts que longs, les textes de moins de 100 mots présentent davantage d'ambiguïtés. Les enfants semblent mieux appréhender les références dans les productions plus longues.