

Le rôle de la structure informationnelle dans l'interprétation d'une anaphore pronominale intraphrastique en français

Saveria Colonna

Laboratoire Structures Formelles du Langage & Université Paris 8
Saveria.Colonna@univ-paris8.fr

Sarah Schimke

Laboratoire Structures Formelles du Langage & Université Paris 8
sarah.schimke@sfl.cnrs.fr

Barbara Hemforth

Laboratoire de Psychologie et de Neuropsychologie Cognitives & Université Paris Descartes
Barbara.Hemforth@univ-paris5.fr

1 Introduction

Pendant la compréhension du langage, le locuteur doit utiliser les marques linguistiques comme les anaphores pour établir les liens entre les différentes entités du discours. Le traitement d'une anaphore implique de trouver un antécédent et le plus souvent, plusieurs antécédents sont disponibles. Une question centrale pour les études sur la compréhension du langage est donc d'une part, de mettre au jour les facteurs qui influencent la résolution anaphorique et d'autre part, de déterminer si le poids de ces facteurs varie suivant la langue du locuteur. Dans cette perspective, nous nous intéressons au cas d'un pronom anaphorique ambigu tel que « il » dans la phrase « Pierre a giflé Jean quand il était jeune ». Alors qu'en allemand, on observe que le pronom ambigu est préférentiellement relié au premier référent mentionné « Pierre » ; en français, il est préférentiellement relié au second référent « Jean » (Colonna, Schimke, Hemforth, Konieczny, & Pynte, 2005). Le même pattern a été observé avec des phrases similaires et la conjonction « avant que » (Hemforth, Colonna, Pynte & Konieczny, 2004). Pour rendre compte de la préférence pour le premier référent, on peut postuler qu'il est préféré parce que c'est le premier mentionné (Gernsbacher, Hargreaves & Beeman, 1989), le sujet de la proposition principale (Crawley, Stevenson & Kleinman, 1990 ; Frederiksen, 1981 ; Järvikivi, van Gompel, Hyönä & Bertram, 2005) et/ou le topique de la phrase (Sanford & Garrod, 1981 ; Crawley, 1986). Mais quels sont les facteurs qui conduisent à préférer le second référent en français ? Les moyens pour marquer la structure informationnelle ne sont pas les mêmes en allemand et en français. En allemand, la structure informationnelle est principalement exprimée en faisant varier l'ordre des mots (et l'accent prosodique), alors qu'en français, l'ordre des mots étant beaucoup plus rigide, on doit utiliser des constructions spécifiques telles que les dislocations ou les clivées. Par conséquent, la première position dans la proposition est beaucoup plus marquée comme étant la position du topique en allemand qu'en français. Est-ce que cette différence peut rendre compte, au moins en partie, de la préférence pour le second référent observée en français ? En effet, même si dans toutes les langues, il est vrai qu'on préfère relier un pronom anaphorique ambigu au topique, il se peut qu'en français, en l'absence de topicalisation explicite, un autre facteur, tel que le rôle grammatical, prenne le pas sur la structure informationnelle et détermine la résolution anaphorique. La préférence pour le second référent serait, dans ce cas, une préférence pour l'objet spécifique au français. Peut-on renverser (ou diminuer) cette préférence pour le second nom en

topicalisant explicitement le premier nom ? Pour répondre à ces questions, nous avons réalisé une série d'expériences off-line sous la forme de trois questionnaires ainsi qu'une expérience on-line avec le paradigme du monde visuel dans lesquelles nous manipulons le statut informationnel et le rôle grammatical du premier référent (N1).

2 Questionnaire 1

2.1 Participants

64 étudiants de l'Université de Genève, de langue maternelle française, ont volontairement complété ce questionnaire.

2.2 Matériel

Nous avons construit 28 phrases expérimentales de la forme suivante : une proposition principale suivie par une proposition subordonnée introduite par « quand ». Le verbe dans la proposition principale était un verbe d'action avec un sujet agent et un objet patient. Tous les sujets et les objets étaient des noms propres. La proposition subordonnée contenait un pronom sujet ambigu qui pouvait se référer soit au sujet, soit à l'objet de la proposition principale. Chaque phrase était suivie d'une phrase dont le début était à compléter par les participants. Les préférences d'interprétation du pronom ambigu étaient ainsi observées, le complètement de la phrase nécessitant l'interprétation du pronom. Un exemple est donné en (1).

(1) N1 sujet et topique :

Pierre a giflé Jean quand il était jeune.

.....était jeune.

Les phrases étaient présentées dans quatre conditions différentes : une condition de base (présentée en 1) où le premier nom (N1) était sujet et topique par défaut, une condition où N1 était objet et topicalisé dans une dislocation à gauche (2), et deux conditions dans lesquelles N1 était focalisé dans une construction clivée et était soit sujet (3), soit objet (4).

(2) N1 objet et topique :

Pierre, Jean l'a giflé quand il était jeune.

(3) N1 sujet et focus :

C'est Pierre qui a giflé Jean quand il était jeune.

(4) N1 objet et focus :

C'est Pierre que Jean a giflé quand il était jeune.

Quatre versions du questionnaire ont été construites de façon à ce que les items apparaissent dans des conditions différentes. Il y avait sept items par conditions dans chaque version du questionnaire. Les items expérimentaux étaient mélangés à quarante items de remplissage de structures syntaxiques variées. Ces items de remplissage étaient eux aussi suivis d'une phrase à compléter mais ne comportaient pas d'ambiguïté référentielle.

2.3 Procédure

Les participants devaient lire chaque phrase puis compléter la phrase qui suivait le plus spontanément possible et sans jamais revenir en arrière. Le complétement du questionnaire prenait une quinzaine de minutes. Les réponses étaient codées suivant que les participants choisissaient le premier ou le deuxième nom mentionné comme antécédent du pronom.

2.4 Résultats et discussion

Les pourcentages d'interprétation du pronom ambigu en faveur du premier nom (N1) dans les quatre conditions testées sont représentés dans la Figure 1. Dans la condition de base (1), le pourcentage d'interprétation du pronom comme se référant à N1 (43,6%) est significativement différent d'une préférence due au hasard ($t_1(63)=2.68, p<.01$; $t_2(27)=2.23, p<.05$). Une analyse de la variance a été réalisée avec le rôle grammatical (sujet vs. objet) et le statut informationnel du premier nom (topique vs. focus) comme variables intra-sujets et les différentes versions du questionnaire comme variable inter-sujets. Elle révèle un effet principal de la structure informationnelle ($F_1(1,60)=21.6, p<.0001$; $F_2(1,27)=13.9, p<.001$) : on relie plus souvent le pronom anaphorique à N1 lorsqu'il est topique que lorsqu'il est focus. On n'observe pas d'effet principal du rôle grammatical ($F_1, F_2<2$). L'interaction entre ces deux facteurs est significative ($F_1(1,60)=4.1, p<.05$; $F_2(1,27)=11.3, p<.01$) : on observe un effet du rôle grammatical en faveur de l'objet seulement dans les conditions où N1 est topique.

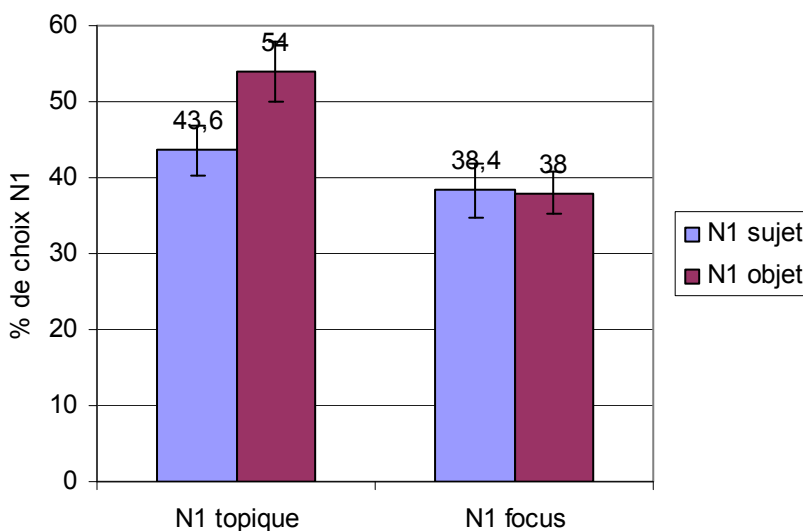


Figure 1 : Pourcentage d'interprétation en faveur de N1 en fonction de son statut grammatical et informationnel.

Un résultat important de ce questionnaire est que, bien que la préférence pour le second référent (N2) déjà observée en français (Hemforth & al., 2004) soit répliquée, cette préférence est nettement diminuée dans la condition (2) où N1 est objet et topicalisé. Il faut cependant noter que dans la condition (2), N1 est objet et explicitement topicalisé par une dislocation à gauche alors que dans la condition de base (1), N1 est sujet et topique par défaut. Il est donc difficile de déterminer si l'augmentation d'interprétation en

faveur de N1 en (2) est due au rôle grammatical (N1 est objet) ou à la structure informationnelle (N1 est explicitement topicalisé). Nous avons donc réalisé un nouveau questionnaire dans lequel nous comparions une condition dans laquelle N1 était sujet et topique explicite (5) à une condition où N1 était objet et topique explicite (6).

(5) N1 sujet et topique explicite :

Pierre, il a vu Jean quand il était jeune.

(6) N1 objet et topique explicite :

Pierre, Jean l'a vu quand il était jeune.

3 Questionnaire 2

3.1 Participants

58 étudiants de l'Université de Paris 8, de langue maternelle française, ont volontairement complété ce questionnaire.

3.2 Matériel

Le matériel expérimental était le même que dans le Questionnaire 1 mais nous n'avions plus que les deux conditions suivantes : N1 était toujours explicitement topicalisé dans une dislocation à gauche mais il était soit sujet (5), soit objet de la principale (6). Nous n'avions donc plus que deux versions du questionnaire.

3.3 Procédure

La procédure était identique à celle utilisée dans le Questionnaire 1.

3.4 Résultats et discussion

Les pourcentages d'interprétation du pronom ambigu en faveur du premier nom (N1) dans les deux conditions testées sont représentés dans la Figure 2. On observe un effet du rôle grammatical ($F(1,56)=2.88, p=.095$; $F(1,27)=5.943, p<.05$). Lorsque N1 est un topique explicite, le pronom ambigu est plus souvent interprété comme se référant à N1 lorsqu'il est objet que lorsqu'il est sujet. De plus, le pourcentage d'interprétation du pronom comme se référant à N1 (56,6%) est marginalement différent d'une préférence due au hasard ($t(57)=1.753, p=.08$; $t(27)=2.454, p=.06$).

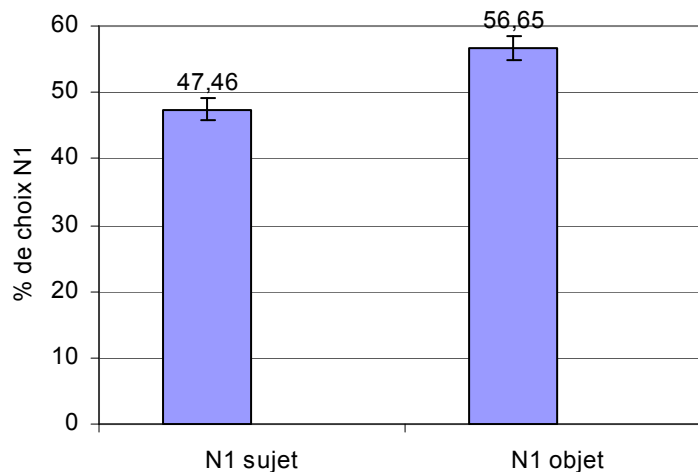


Figure 2 : Pourcentage d'interprétation en faveur de N1 en fonction de son rôle grammatical.

Ces résultats montrent qu'à statut informationnel égal, les locuteurs du français préfèrent interpréter le pronom ambigu comme se référant à l'objet plutôt qu'au sujet de la principale. Ils éclairent par ailleurs la préférence pour N2 observée dans la condition de base (1) qui pouvait être interprétée comme une préférence pour l'objet mais aussi comme une préférence pour le second référent mentionné. Or, nous avons vu dans ce questionnaire, que l'objet est préféré même lorsqu'il est le premier référent mentionné (6). On peut cependant objecter que les structures disloquées que nous avons utilisées, même si elles sont fréquentes en français oral, sont beaucoup plus rares à l'écrit. Nous avons donc répliqué cette manipulation avec des structures plus fréquentes à l'écrit dans le Questionnaire 3.

4 Questionnaire 3

4.1 Participants

51 étudiants de l'Université de Paris 8, de langue maternelle française, ont volontairement complété ce questionnaire.

4.2 Matériel

Le matériel expérimental était le même que dans les questionnaires précédents mais dans les conditions suivantes : une condition de base (1) où N1 était sujet et topique par défaut (identique à celle du questionnaire 1), et deux conditions dans lesquelles N1 était explicitement topicalisé par la construction « quant à » et était soit sujet (2), soit objet (3).

(1) N1 sujet et topique par défaut :

Pierre a giflé Jean quand il était jeune.

(2) N1 sujet et topique explicite :

Quant à Pierre, il a giflé Jean quand il était jeune.

(3) N1 objet et topique explicite :

Quant à Pierre, Jean l'a giflé quand il était jeune.

Nous n'avions plus que 27 phrases expérimentales et trois versions du questionnaire.

4.3 Procédure

La procédure était identique à celle utilisée dans le Questionnaire 1.

4.4 Résultats et discussion

Les pourcentages d'interprétation du pronom ambigu en faveur du premier nom (N1) dans les trois conditions testées sont représentés dans la Figure 3. Comme dans le Questionnaire 1, dans la condition de base (1), le pourcentage d'interprétation du pronom comme se référant à N1 (38,5%) est significativement différent d'une préférence due au hasard ($t_1(50)=2.97$, $p=.005$; $t_2(26)=4.122$, $p=.0001$). Une analyse de la variance a été réalisée pour comparer les conditions (2) et (3) avec le rôle grammatical comme variable intra-sujets et les différentes versions du questionnaire comme variable inter-sujets. Elle révèle un effet principal du rôle grammatical ($F_1(1,48)=3.569$, $p=.065$; $F_2(1,26)=4.354$, $p<.05$) : on relie plus souvent le pronom anaphorique à N1 lorsqu'il est objet que lorsqu'il est sujet. De même, si on compare les conditions (1) et (2) et les conditions (1) et (3), on s'aperçoit qu'on interprète plus souvent le pronom comme se référant à N1 lorsque celui-ci est explicitement topicalisé que lorsqu'il est topique par défaut (respectivement, $F_1(1,48)=2.229$, $p>.10$; $F_2(1,26)=3.556$, $p=.071$ et $F_1(1,48)=7.822$, $p<.01$; $F_2(1,26)=12.585$, $p<.01$).

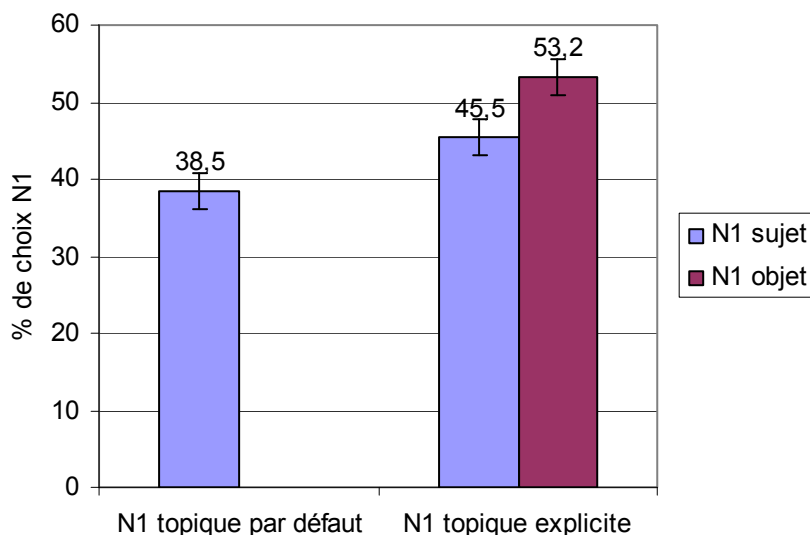


Figure 3 : Pourcentage d'interprétation en faveur de N1 en fonction de son statut grammatical et informationnel.

5 Discussion des expériences off-line

L'ensemble des résultats des trois questionnaires révèle qu'on préfère relier un pronom intra-phrastique ambigu au topique plutôt qu'au focus de la phrase (Questionnaire 1), et que cette préférence est plus forte lorsqu'il s'agit d'un topique explicite que lorsqu'il s'agit d'un topique par défaut (Questionnaire 3). Les trois questionnaires révèlent par ailleurs un effet du rôle grammatical : on préfère relier le pronom à l'objet plutôt qu'au sujet de la proposition principale. Cette préférence pour l'objet n'a été observée qu'en français et en portugais (Baumann, Konieczny & Hemforth, 2010), les autres langues étudiées révélant une préférence pour le sujet dans des structures comparables (Hemforth & al., 2004 ; Colonna, Schimke, Hemforth & Escobar, 2009). Nous reviendrons sur cette différence inter-langue dans la discussion générale.

Les préférences que nous avons observées dans ces questionnaires reflètent l'interprétation finale des locuteurs mais ne nous renseignent pas sur les processus de résolution anaphoriques en temps réel. Nous avons donc réalisé une expérience avec le paradigme du monde visuel afin de pouvoir observer les préférences d'interprétation du pronom anaphorique en temps réel. Ce paradigme consiste à faire écouter au participant un énoncé oral faisant référence à certains éléments représentés dans une scène visuelle pendant qu'on enregistre ses mouvements oculaires. Les préférences d'interprétation du pronom ambigu peuvent ainsi être inférées à partir du pattern d'exploration visuelle développé sur la scène visuelle représentant les différents référents possibles (voir Figure 4).



Figure 4 : Exemple de scène visuelle.

6 Expérience en temps réel

6.1 Participants

32 étudiants de l'Université de Paris 8, de langue maternelle française, ont volontairement participé à cette expérience.

6.2 Matériel

Nous avons construit vingt items expérimentaux et vingt items de remplissage à partir du matériel utilisé pour les questionnaires. Les noms propres ont été remplacés de façon à rendre les référents visuellement identifiables. De plus, nous avons utilisé le futur afin que le nombre du verbe dans la proposition subordonnée soit marqué phonologiquement et qu'ainsi, le pronom « il » ne puisse pas être pris pour un pronom pluriel se référant aux deux antécédents. Par exemple, à la Figure 4 correspondait l'énoncé oral : « Voici le facteur et le balayeur. Quant au facteur, il va gifler le balayeur quand il sera à la maison » dans une des quatre conditions suivantes : N1 était sujet et topique explicite (1), objet et topique explicite (2), sujet et focus (3) ou objet et focus (4).

(1) N1 sujet et topique :

Voici le facteur et le balayeur. Quant au facteur, il va gifler le balayeur quand il sera à la maison.

(2) N1 objet et topique :

Voici le facteur et le balayeur. Quant au facteur, le balayeur va le gifler quand il sera à la maison.

(3) N1 sujet et focus :

Voici le facteur et le balayeur. C'est le facteur qui va gifler le balayeur quand il sera à la maison.

(4) N1 objet et focus :

Voici le facteur et le balayeur. C'est le facteur que le balayeur va gifler quand il sera à la maison.

Les images étaient toutes similaires à celle présentée dans la Figure 4 et comportaient deux personnages. Afin de contrôler d'éventuelles préférences de balayage oculaires, le premier référent mentionné était soit présenté à droite, soit à gauche de l'écran.

6.3 Procédure et dispositif expérimental

Les mouvements oculaires étaient enregistrés par un oculomètre Eyelink 1000 (SR research). Après que le participant ait fixé un point au centre de l'écran, l'image apparaissait à l'écran et le fichier son associé se déclenchait. Les participants devaient ensuite juger si l'image correspondait ou pas à ce qu'ils avaient entendu. Pour la moitié des items de remplissage, les images ne correspondaient pas à la phrase orale entendue.

6.4 Résultats et discussion

La figure 5 présente la probabilité de regarder le premier référent (plutôt que le second) à partir du moment où le participant entend le pronom ambigu et pendant les 2 secondes qui suivent. Des analyses de la variance ont été réalisées pour chaque fenêtre de temps de 400 millisecondes. Ces analyses révèlent un effet du rôle grammatical, N1 a plus de probabilité d'être fixé lorsqu'il est objet que lorsqu'il est sujet (deuxième fenêtre : $F(1,31)=5.73$, $p<.05$; $F(1,19)=10.46$, $p<.01$; troisième fenêtre : $F(1,31)=17.8$, $p<.001$; $F(1,19)=26.37$, $p<.001$; quatrième fenêtre : $F(1,31)=12.75$, $p<.01$; $F(1,19)=18.3$, $p<.001$) ; ainsi qu'un effet du rôle informationnel, N1 a plus de probabilité d'être fixé lorsqu'il est topicalisé que lorsqu'il est focalisé (deuxième fenêtre : $F(1,31)=5.35$, $p<.05$, $F(1,19)=6.92$, $p<.05$; troisième fenêtre : $F(1,31)=8.56$, $p<.01$; $F(1,19)=8$, $p<.01$; quatrième fenêtre : $F(1,31)=3.7$, $p=.064$; $F(1,19)=3.74$, $p=.068$). L'interaction n'est pas significative ($F_s<2$). Ces résultats confirment le rôle de la structure informationnelle observé dans les questionnaires : on relie préférentiellement le pronom ambigu au nom

topicalisé plutôt qu'au nom focalisé. De plus, ils révèlent un effet en temps réel du rôle grammatical en faveur de l'objet et ce, même dans la condition focus. Ce dernier résultat montre que les locuteurs du français relient préférentiellement le pronom anaphorique à l'objet et non simplement au référent le plus récent.

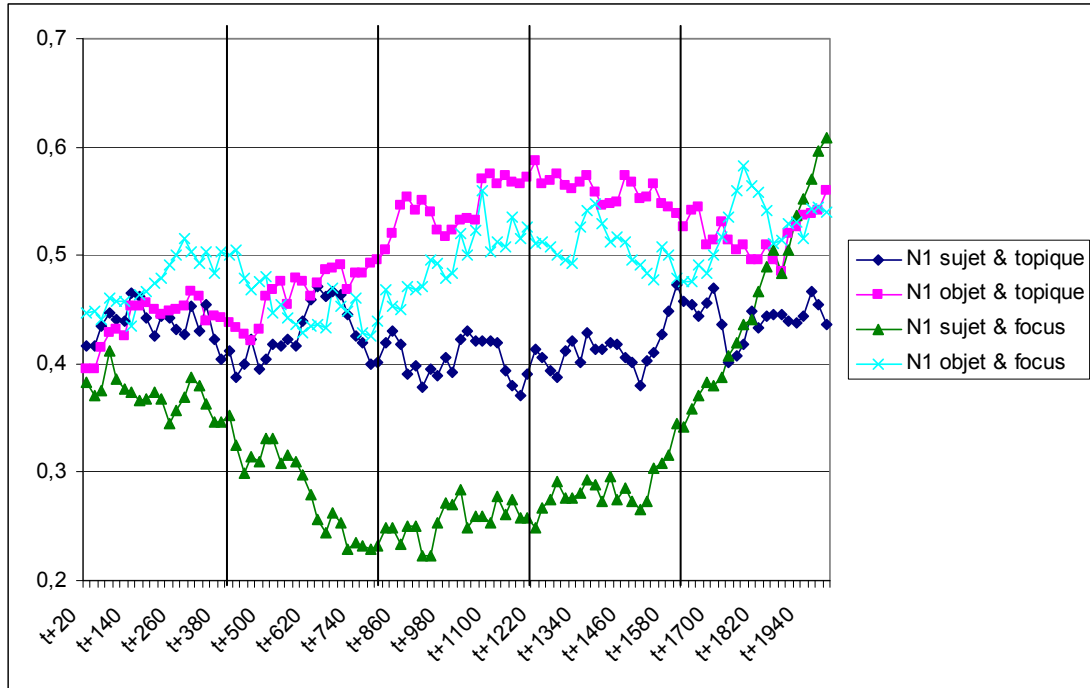


Figure 5 : Probabilité de regarder le premier référent mentionné (N1)

7 Discussion générale

L'ensemble de nos résultats révèle que la structure informationnelle influence le choix de l'antécédent, on préfère un référent topicalisé à un référent focalisé. Cette préférence ne peut s'expliquer en termes de saillance puisqu'on peut soutenir que les éléments topicalisés comme les éléments focalisés sont saillants dans le discours (Cowles, Walenski & Kluender, 2007). L'idée suivant laquelle le premier référent mentionné serait le référent préféré (Gernsbacher & al., 1989) ne peut pas non plus expliquer ce résultat. En effet, les référents que nous avons soit topicalisés, soit focalisés, occupaient toujours la première position. Ce sont les fonctions discursives différentes de la topicalisation et de la focalisation qui permettent d'expliquer pourquoi le pronom dans la proposition subordonnée est préférentiellement relié au référent marqué comme topique plutôt qu'au référent marqué comme focus. En effet, interpréter le pronom comme se référant au topique contribuerait à la cohérence du discours en évitant de changer de topique de la phrase entre la proposition principale et la proposition subordonnée.

D'autre part, nos résultats révèlent un effet du rôle grammatical : on préfère relier le pronom anaphorique à l'objet plutôt qu'au sujet de la principale. Cette préférence pour l'objet semble contribuer pour beaucoup à la préférence observée pour le second nom dans la condition de base (1). En effet, en l'absence de marqueurs de topicalisation explicites, les locuteurs du français semblent préférentiellement

relier le pronom à l'objet de la principale. Une explication possible s'appuie sur le fait qu'en français, on relie fréquemment et sans ambiguïté une proposition subordonnée au sujet de la principale en utilisant une forme infinitive telle que « Pierre a frappé Jean avant de rentrer à la maison ». Conséquemment, suivant un principe Griceen, les locuteurs du français peuvent dans des cas ambigus relier le pronom anaphorique préférentiellement à l'objet. Il se peut que les locuteurs généralisent cette préférence pour l'objet aux propositions subordonnées introduites par « quand » pour lesquelles la forme infinitive n'est pas possible (mais pour lesquelles il est cependant possible d'utiliser une forme non ambiguë avec un gérondif telle que « Pierre a frappé Jean étant jeune »). Un premier argument en faveur de cette hypothèse est qu'en allemand, où une préférence claire pour le sujet a été observée (Colonna & al., 2005 ; Hemforth & al., 2004), une forme alternative non ambiguë n'existe pas. Au contraire, en portugais, où comme en français, une forme non ambiguë existe, une préférence pour l'objet a été observée (Baumann & al., 2010). D'autre part, en anglais où il existe aussi une forme alternative non ambiguë pour se référer au sujet de la principale (« *Peter slapped John before going home* »), une préférence pour le sujet a néanmoins été observée. Cependant, cette préférence était beaucoup moins forte que celle observée en allemand (Hemforth, Scheepers, Colonna, Pynte, & Konieczny, 2005). De plus, il faut noter que la forme alternative anglaise « *before -ing* » est beaucoup moins fréquente que la forme infinitive française (Hemforth, Scheepers, Konieczny, Pynte & Colonna, 2009). Ceci peut expliquer pourquoi en anglais, la préférence pour le sujet diminue (par rapport à ce qui a été observé en allemand) mais n'est pas renversée (comme c'est le cas en français et en portugais). D'autres comparaisons inter-langues, en fonction de l'existence (et de la fréquence) de formes alternatives non ambiguës, doivent être réalisées pour tester cette hypothèse.

Références bibliographiques

- Baumann, P., Konieczny, L. & Hemforth, B. (2010). Expecting coreference: the role of alternative constructions. *CUNY Conference on Human Sentence Processing*, New York, March 16-20.
- Colonna, S., Schimke, S., Hemforth, B. & Escobar M. (2009). Language specific preferences in pronoun resolution: Evidence from French, Spanish and German. *AMLaP Conference*, Barcelona, September 7-9.
- Colonna, S., Schimke, S., Hemforth, B., Konieczny, L & Pynte, J. (2005). The role of information structure in anaphora resolution. *AMLaP Conference*, Ghent, September 5-7.
- Cowles, H.W., Walenski, M., & Kluender, R. (2007). Linguistic and cognitive prominence in anaphor resolution: topic, contrastive focus and pronouns. *Topoi*, 26,3-18.
- Crawley, R. A. (1986). Some factors influencing the comprehension of pronouns. In Clifton, C. (Ed.), *Proceedings of the Eighth Annual Conference of the Cognitive Science Society* (pp. 613-620). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crawley, R., Stevenson, R., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, 4, 245-264.
- Frederiksen, J. (1981). Understanding anaphora: Rules used by readers in assigning pronominal referents. *Discourse Processes*, 4, 323-347.
- Gernsbacher, M.A., Hargreaves, D.J., & Beeman, M. (1989). Building and accessing clausal representations: The advantage of first mention versus the advantage of clause recency. *Journal of Memory and Language*, 28, 735-755.
- Hemforth, B., Colonna, S., Pynte, J. & Konieczny, L. (2004). Pronoun resolution across languages. Architectures and Mechanisms for Language Processing Conference, Aix en Provence, September 16-17.

- Hemforth, B., Scheepers, C., Colonna, S., Pynte, J. & Konieczny, L. (2005). Pronoun resolution across languages: The role of distributional differences between French, German, and English. *CUNY*, Tucson, March 31-April 2.
- Hemforth, B., Scheepers, C., Konieczny, L., Pynte, J., & Colonna, S. (2009). Language specific preferences in anaphor resolution. *Text and Discourse*, Rotterdam, July 26-28.
- Järvikivi, J., van Gompel, R.P.G, Hyönä J. & Bertram R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject preference accounts. *Psychological Science*, 16, 260-264.
- Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language. Explorations in comprehension beyond the sentence*. Chichester: J. Wiley and Sons.