Prof. Dr. Sandrine Zufferey

UNIVERSITÄT BERN

Etudier le marquage des relations de cohérence : corpus et expériences

LINGUISTIQUE DE CORPUS – SÉMINAIRE MA
UNIVERSITÉ DE NEUCHÂTEL

Objectifs du cours

- Comprendre l'utilité des **corpus parallèles** pour répondre à différents types de questions de recherche.
- Découvrir la complémentarité des **approches empiriques** par le biais de **corpus** et d'**expériences**.
- Réfléchir au **rôle des connecteurs** dans le **marquage** des **relations de cohérence**.

Le marquage linguistique des relations de cohérence

Discours et relations de cohérence

- Dans un discours, les énoncés **ne se suivent pas aléatoirement**.
- Entre **chaque segment** d'un discours, il existe une **relation de cohérence** qui reflète **l'articulation des idées**.
- **Quelques exemples** de relations :
  - *Marie est malade, **donc** elle n'ira pas à l'école aujourd'hui.*
  - *La porte a claqué, **parce qu**'il y a un fort vent dehors.*
  - *Jean a mis son manteau, **puis** il est parti à l'école.*
  - *Pierre est grand, **mais** Alfred est petit.*

Quand doit-on utiliser un connecteur ?

Livre pour enfants

L'été avait été très beau. Ce jour là, dans le ciel, des nuages noirs montaient. Ce fut la nuit, l'eau tomba à torrents. Tous se dépêchaient de trouver un abri, la cigale rentra chez elle, elle avait des travaux ménagers à faire.

Article de journal

Nos océans sont malades. **En effet**, 80% des stocks de poissons sont surexploités. **Par ailleurs**, les ressources de la mer s'épuisent. **En outre**, nous avons trop de bateaux, **mais** plus assez de poissons **et** aucun respect pour la mer. **D'un côté**, nous lui prenons ce qui nous intéresse **et d'un autre côté** nous y laissons ce dont nous n'avons plus besoin.

5

Différence entre les relations de cohérence

1) a. Jeanne est contente **parce qu**'elle a gagné la course.
   b. Jeanne est contente. Elle a gagné la course. ✓
2) a. Jeanne est contente **même si** elle n'a pas gagné la course.
   b. Jeanne est contente. Elle n'a pas gagné la course. ✗
3) a. Max a pris son petit déjeuner. **Après,** il est parti à l'école.
   b. Max a pris son petit déjeuner. Il est parti à l'école. ✓
4) a. Max est parti à l'école. **Avant,** il a pris son petit déjeuner.
   b. Max est parti à l'école. Il a pris son petit déjeuner. ✗

6

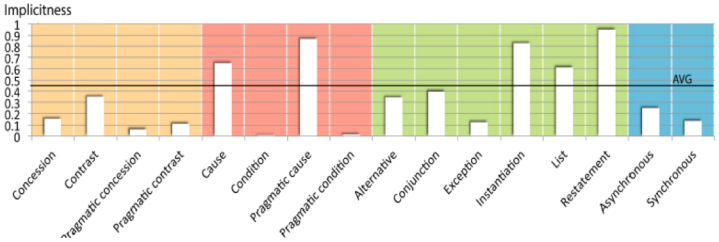
La continuité : une attente cognitive

- Les lecteurs s'attendent à ce que les relations soient **temporellement continues** et/ou **causales** (Murray, 1997; Sanders, 2005).
- Les **relations continues** sont facilement **inférables** sans connecteur.
  - ✓ Max a pris son petit déjeuner. Il est parti à l'école. (**continuité temporelle**)
  - ✓ Jeanne est contente. Elle a gagné la course. (**continuité causale**)
- Les **relations discontinues** nécessitent un **connecteur explicite**.
  - ✗ Max est parti à l'école. Il a pris son petit déjeuner. (**inversion temporelle**)
  - ✗ Jeanne est contente. Elle n'a pas gagné la course. (**négation de la cause**)

7

Etude sur le corpus de la PDTB

- Calcul du ratio d'implicite pour chaque relation de cohérence.
  - nb d'occurrences implicites / nb total d'occurrences d'une relation



Asr & Demberg, 2012 : 10

8

Questions ouvertes

- Certaines relations qui ne sont liées **ni** à la **causalité ni** à la **temporalité** sont très souvent **traduites explicitement**. **Pourquoi** ?
- L'étude de la PDTB ne porte que sur **l'anglais**. Ces différences sont-elles **universelles** ?
- Le **taux d'accord** entre les annotateurs est plus **bas** pour les **relations implicites**. Comment **contourner** ce **problème** ?

9

Hypothèses

- La **continuité** est une **contrainte discursive très générale**.
  - continuité d'espaces mentaux (relation de condition)
  - **continuité de perspective narrative** (relations de correction et de confirmation)
- L'attente de continuité est une **contrainte cognitive.**
  - effets visibles pendant la **lecture**
- C'est une **contrainte universelle** au travers des langues.
  - vérification dans des **corpus multilingues**

10

Le rôle des changements de perspective narrative

- Dans un texte, les informations sont présentées à partir d'une certaine **perspective narrative** (Sanders, 1994 ; Pander Maat, 1998).
  - subjective interne : *Je pense que Pierre est coiffeur.*
  - subjective externe : *Emma pense que Pierre est coiffeur.*
  - objective (acceptée par le locuteur) : *Pierre est coiffeur.*
- Certaines relations comme la **correction** et la **confirmation** impliquent un **changement de perspective narrative** entre les deux segments :
  - *Emma pense que Pierre est coiffeur. En fait, il travaille comme vendeur.*
  - *Emma pense que Pierre est coiffeur. Et en effet, c'est bien son métier.*

11

Traductions des relations de cohérence dans des corpus multilingues

La traduction des connecteurs

- **Constat**
  - les connecteurs sont **volatiles** dans les traductions
  - fréquents **ajouts** et **omissions** (Halverson, 2004 ; Zufferey & Cartoni, 2014)
- **Hypothèse : ces variations ne sont pas aléatoires**
  - les connecteurs qui **signalent un relation continue** sont **plus souvent traduits par une relation implicite** que ceux qui indiquent une relation discontinue.
  - cette différence est **constante entre les langues**, quelles que soient les équivalents de traduction présents dans la langue cible.

Le cas de *en effet*

- Relation de **cause : continuité narrative**
  - *Emma ne sera bientôt plus célibataire. En effet, elle va épouser Max en juin.*
  - *Emma ne sera bientôt plus célibataire. Elle va épouser Max en juin.* ✓
- Relation de **confirmation : discontinuité narrative**
  - *Jeanne craignait que Max n'épouse Emma. Et en effet, ils se sont mariés en juin.*
  - *Jeanne craignait que Max n'épouse Emma. Ils se sont mariés en juin.* ✗
- **Hypothèse**
  - les **relations de cause** transmises par *en effet* sont **plus souvent traduites implicitement** que les relations de confirmation

Le corpus Europarl

- Débats du **parlement européen** (Koehn, 2005).
  - les députés parlent dans leur langue maternelle
  - puis les débats sont retranscrits et traduits dans autres langues officielles
- Corpus parallèles directionnels (Cartoni, Zufferey & Meyer, 2013).

français langue source → **allemand** langue cible
anglais langue cible
espagnol langue cible

Méthode: le translation spotting

- **français** langue source: **extraction** de 500 *en effet*, **annotation** de la relation
- **allemand**, **anglais**, **espagnol** langues cibles: identification de la traduction

langue source	langue cible
Les autorités américaines doivent s'y attendre; elles sont **en effet** tout à fait conscientes de l'opposition de l'Unio[illegible]mble à la peine de mort et à toute exécution. (cause)	Diese sind im übrigen darauf gefaßt, **denn** sie wissen recht genau, daß die Union in ihrer Gesamtheit die Todesstrafe und jede Hinrichtung ablehnt.
La résolution ne dit rien, **en effet**, sur le problème essentiel, [illegible] la politique menée dans l'ex-Yougoslavie. (confirmation)	The resolution **indeed** is silent on the key issue, which is the validity of the policy implemented in the former Yugoslavia.
Nous aimerions **en effet** que la Commission nous donne plus de dét[illegible] bonne application de ce prin[illegible] proposition, afin de dresser un véritable bilan lors de son rapport annuel. (cause)	Nos gustaría **Ø** que la Comisión nos diera más detalles sobre la buena aplicación de este principio en cada propuesta, para hacer un verdadero balance en su informe anual.

Résultats : allemand langue cible

	zéro	*denn*	*nämlich*	*in der Tat*	autre	Total
Cause	153	75	57	32	103	422
Confirmation	10	6	4	28	32	78
Total	163	81	61	60	135	500

	explicite	implicite
Cause	64%	36%
Confirmation	88%	12%

- Khi carré d'indépendance : $\chi^2 = 6.95$, df = 1, $p < 0.01$.
- Conclusion : les **relations de confirmation** sont **traduites explicitement plus souvent** que les relations causales.

Résultats : anglais langue cible

	zéro	*indeed*	*in fact*	autre	Total
Cause	202	51	75	94	422
Confirmation	19	35	5	19	78
Total	221	86	80	113	500

	explicite	implicite
Cause	52%	48%
Confirmation	76%	24%

- Khi carré d'indépendance : $\chi^2 = 8.33$, df = 1, $p < 0.01$.
- Conclusion : les **relations de confirmation** sont **traduites explicitement plus souvent** que les relations causales.

Résultats: espagnol langue cible

	en efecto	*efectivamente*	zero	autre	Total
Cause	237	64	76	14	391
Confirmation	64	28	12	5	109
Total	301	92	88	19	500

	explicite	implicite
Cause	81%	19%
Confirmation	89%	11%

- Khi carré d'indépendance : $\chi^2 = 4.2$, df = 1, $p < 0.05$.
- Conclusion : les **relations de confirmation** sont **traduites explicitement plus souvent** que les relations causales.

Conclusions de l'étude de corpus

- La proportion de traductions implicites **varie en fonction des équivalents de traduction** disponibles dans la langue cible.

MAIS

- Les traductions implicites ne sont **pas aléatoires** : elles sont liées aux **relations causales**, qui sont discursivement **continues**.
- L'hypothèse du rôle de la **continuité de perspective narrative** est **renforcée** par les données de corpus multilingues.

Traitement des relations de cohérence durant la lecture

Différences cognitives entre explicite et implicite

- La **lecture** du segment textuel qui suit un connecteur (explicite) est **plus rapide** qu'en l'absence de connecteur (implicite).
  - Britton et al., 1982 ; Sanders & Noordman, 2000 ; Zufferey, 2014
- Les lecteurs **comprennent mieux** les relations explicites.
  - tests de questions de compréhension portant sur le texte
  - Millis & Just, 1994 ; Degand & Sanders, 2002 ; van Silfhout et al., 2014
- La **communication implicite** d'une relation produit des résultats différents selon leur caractère continu ou discontinu.
  - enlever le connecteur d'une relation discontinue perturbe le traitement
  - Murray, 1995; 1997

Expérience de lecture

- Participants: **31 locuteurs** natifs du français
- Matériel: **40 phrases** test en **4 versions** différentes

	cause	**confirmation**
explicite	Albert a dû faire une grosse bêtise hier soir. En effet, **son père a confisqué son vélo pour un mois.**	Albert pensait qu'il serait puni pour avoir poussé sa sœur. Et en effet, **son père a confisqué son vélo pour un mois.**
implicite	Albert a dû faire une grosse bêtise hier soir. **Son père a confisqué son vélo pour un mois.**	Albert pensait qu'il serait puni pour avoir poussé sa sœur. **Son père a confisqué son vélo pour un mois.**

Exemple d'auto-présentation segmentée

_____ __ _____ __ _____ segmentée.

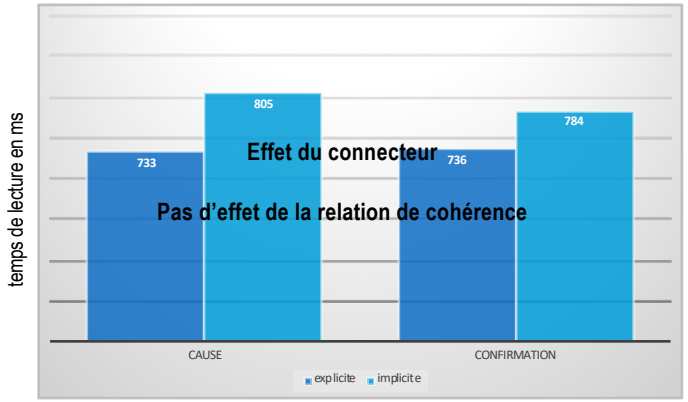
Expérience de lecture en ligne : procédure

- Division de la phrase en **7 segments de lecture.**
- Un **exemple**
  [Albert pensait 1] [qu'il serait puni 2] [pour avoir poussé sa sœur. 3]
  [Et en effet, 4] **[son père a confisqué 5] [son vélo 6] [pour un mois. 7]**
- **Analyse** des données
  – temps de lecture en millisecondes des segments 5, 6 et 7

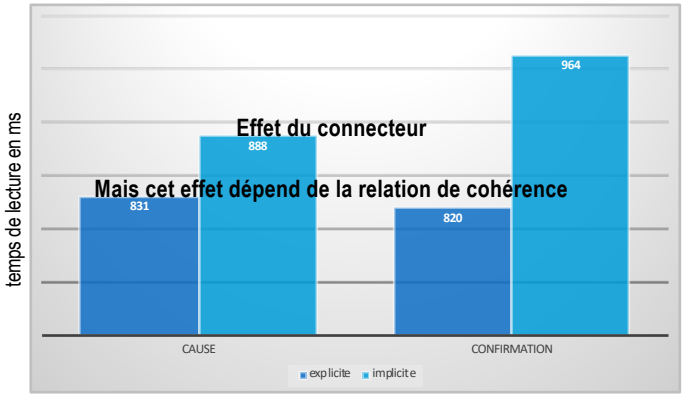
Expérience de lecture : hypothèses

- **Effet du connecteur**
  – **accélération** de la lecture en **région 5** pour **toutes** les **relations explicites**
- **Effet de la relation de cohérence**
  – **ralentissement** en **région 7** pour les **relations de confirmation implicites** mais pas pour les **relations de cause implicites**

Résultats I: juste après le connecteur (région 5)



Résultats II: à la fin de la phrase (région 7)



Conclusions de l'expérience de lecture

- Cette expérience **confirme l'accélération immédiate** de la lecture provoquée par les connecteurs.
  - réplication d'un effet connu dans la littérature
- Cette expérience **démontre le rôle de la continuité narrative** pour le traitement des relations de cohérence.
  - les relations qui impliquent un **changement de perspective narrative** comme la confirmation **doivent être marquées explicitement**

29

En résumé

- Les relations de cohérence peuvent être rendues **explicites** par un connecteur ou communiquées **implicitement**.
  - grandes différences entre les types de relations
- Les relations de cohérence qui sont **temporellement**, **causalement** ou **narrativement continues** sont facilement **inférables**.
  - elles sont souvent communiquées implicitement dans toutes les langues
- **Ne pas marquer explicitement** des relations **discontinues** crée des **problèmes** de lecture et de compréhension.

30

Références

- Zufferey, S. (2016). Discourse connectives across languages. Factors influencing their explicit or implicit translation. *Languages in Contrast* 16(2), 264–279.
- Zufferey S. & Gygax P. (2016). Discourse relations with a perspective shift are difficult to convey implicitly. Evidence from processing and translating the French connective 'en effet'. *Discourse Processes* 53(7), 532–555.

31