Repérage automatique de la relation de contraste

Approches par marqueurs lexico-syntaxiques, antonymie et similarité structurelle

LALLEMAN Fanny
MORLANE-HONDERE François
TULECHKI Nikola

20/01/2009

1. Introduction
   1. La relation de contraste
   2. TAL et Discours
      1. Enjeux, Méthodes et Limites
      2. Principales Applications
2. Repérage automatique de contraste
   1. Approche par marqueurs
   2. Approche par antonymie
   3. Approche par parallélisme syntaxique
3. Conclusion et perspectives

La relation de contraste

- *Relation interne qui exprime une opposition entre deux propositions* (Busquets, 2007)
- Un grand nombre de théories se sont intéressées à cette relation de contraste (Sporleder&Lascarides, 2006) :
  - Rhetorical Structure Theory (RST)
  - Discourse Representation Theory (DRT)
  - Segmented Discourse Representation Theory (SDRT)
  - Discourse Lexicalised Tree Adjoining Grammar (DLTAG)

But ...

- Multitude de typage de la relation de contraste (Busquets, 2007) :

Hobbs(198	RST	SDRT	Autres
Contraste Violation d'attente	Contraste Concession Antithèse Anti-condition	Contraste	Opposition sémantique Violation d'attente Concession

  - La plupart de ces classifications se sont basées sur le connecteur *but* pour définir plusieurs types de contraste, en fonction de sa distribution et de son fonctionnement.

Différents types d'expression du contraste (Busquets, 2007)

- Opposition sémantique (*semantic opposition*)
  - Similarité syntaxique et antonymie
- Violation d'attente (*denial of expectation*)
  - Relation de contraste indirecte
- Antithèse (*antithesis*) et anti-condition (*otherwise*)
  - Contraste à valeur argumentative et intentionnelle

Opposition sémantique

- *Antonymie entre prédicats qui sont comparables en un certain sens* (Lakoff, 1971)
  - Axe commun autour duquel le contraste est possible (Widlöcher, 2008)

En attendant de trouver une autre manière de généraliser à partir de « contextes » par définition changeants (Bensa 1996 :44), c'est bien sous forme de système non pas intemporels, mais simplement stables dans la moyenne durée(…) [CHP-2820]

Violation d’attente

- Contraste ou opposition entre énoncés :
  - Nécessité de présupposer un monde connu à partir duquel on peut inférer un contraste.
  - Contraste de nature indirecte

C'est dans les années 1880 qu'une littérature sur les abus sexuels des enfants fit son apparition, **bien que** le terme de pédophile n'émergea qu'en 1925. [340_CHP]

Contraste à valeur argumentative et intentionnelle

- Antithèse :
  - Incompatibilité entre propositions, le locuteur portant une préférence à la situation exprimée dans le noyau (Busquets, 2007).
- Anti-condition :
  - La réalisation de la proposition principale empêche celle de la subordonnée (Busquets, 2007).
- Concession :
  - Echec par rapport à un évènement attendu ou espéré.

Différences et opposition (Busquets, 2007)

Contraste/Opposition sémantique

- Marqueurs non obligatoires pour signaler la relation
- N'implique pas la concession
- Propositions équivalentes
- Pas d'attentes inférées
- Additive
- Relation symétrique

Concession et contraste à valeur argumentative

- Essentiellement signalé par marqueurs
- Implique une opposition indirecte
- Une proposition proéminente
- Attentes inférées
- Causale
- Relation asymétrique

1. Introduction
   1. La relation de contraste
   2. **TAL et Discours**
      1. Enjeux, Méthodes et Limites
      2. Principales Applications
2. Repérage automatique de contraste
   1. Approche par marqueurs
   2. Approche par antonymie
   3. Approche par parallélisme syntaxique
3. Conclusion et perspectives

"It is likely that the field of discourse parsing will have a quantifiable impact on many end-to-end applications, not immediately, but in a slightly more distant future. However, as adequate solutions to the apparently simpler problems are found and as progress is made in discourse linguistics, the field of discourse parsing is poised to drive significant advances in natural language and enable many new applications that we cannot even conceive of today."

D. MARCU

- Analyse automatique du discours
  - Discipline récente
  - *Dériver* la *structure discursive* en *inférant* les *relations du discours* qui relient les *unités du discours* entre elles.

    - *Quelle représentation?*
    - *Quelles relations?*
    - *Quelles unités?*
    - *Quels observables?*

- Représentations
  - Séquences
    - Séquences de segments de continuité topicale
  - Arbres
    - Unités minimales progressivement regroupées dans des unités d'ordre supérieure
  - Graphes
    - Peu contraignants mais difficilement validables d'un point de vue linguistique

- Relations
  - Désaccord entre les théories sur le nombre et les caractéristiques des relations.
  - Définition de « relation de discours » varie et dépend souvent des moyens précis utilisés pour l'établir.
  - Le jeu de relations peut dépendre de l'application envisagée.

- Unités
  - Aucun consensus sur ce que c'est une unité de discours
    - Proposition (?)
    - Phrase typographique
    - Paragraphe

- Observables
  - Méthodes d'analyse inadaptées pour l'analyse automatique
    - Critères pragmatiques (intention du locuteur)
    - Connaissance du domaine
    - Inférences
    - *Compréhension du message*
  - Principaux observables
    - Marqueurs (adverbes, prépositions, locutions adverbiales…)
    - Cohésion lexicale
    - Syntaxe
    - Expressions référentielles
    - Sémantique

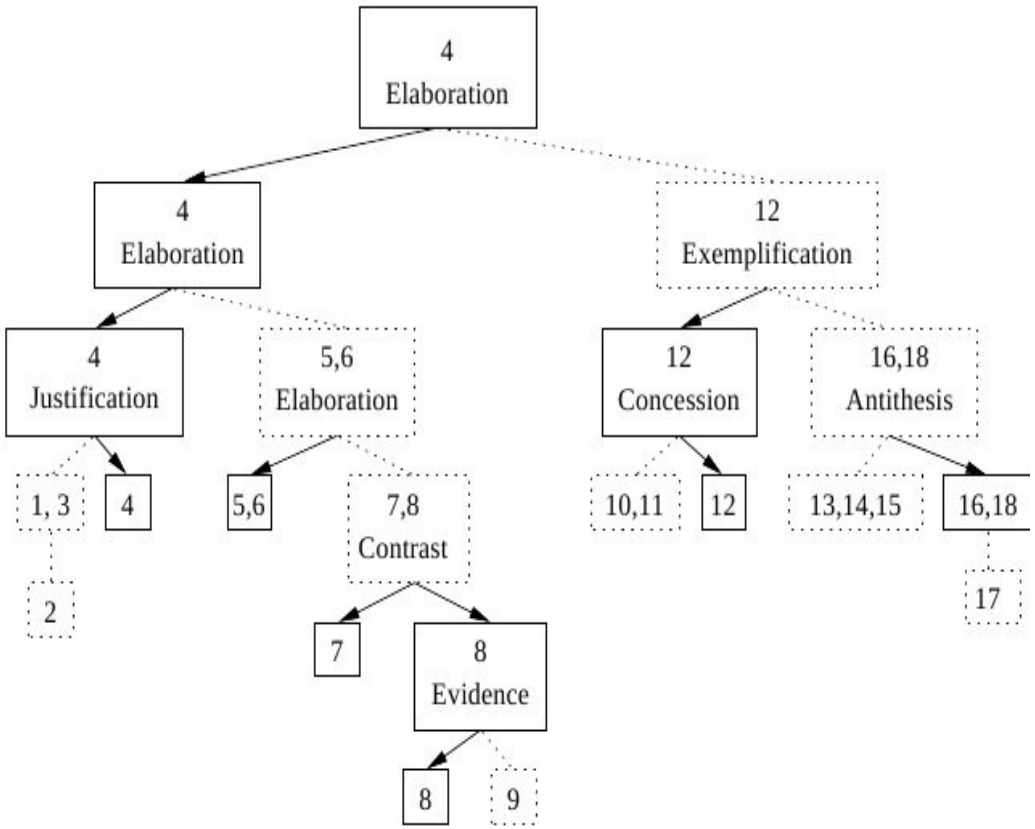
Principales applications du repérage automatique des relations de discours

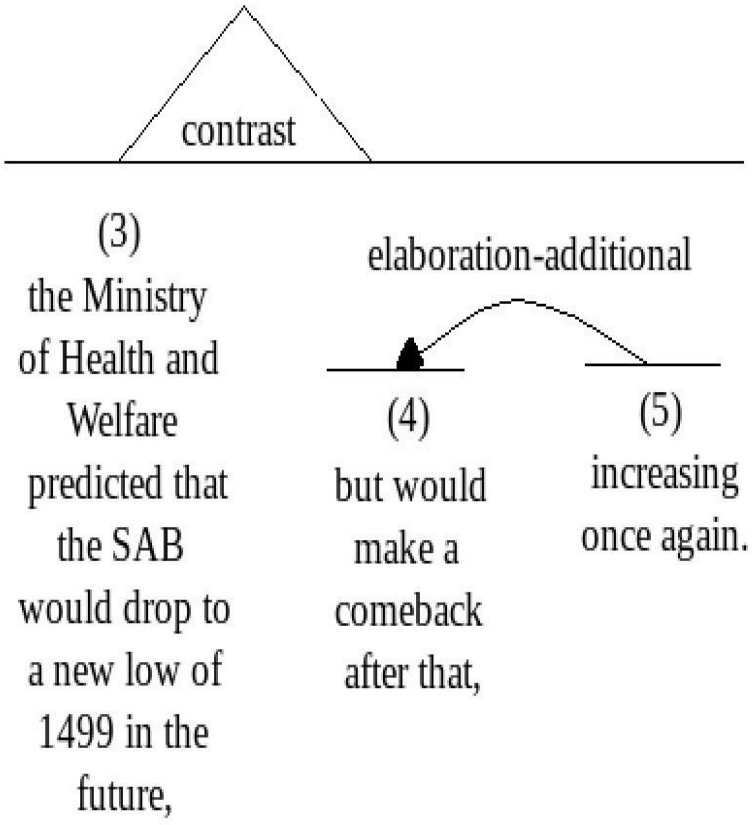
- Applications basées sur le modèle de la RST
  - En résumé automatique
  - En recherche d'information
  - En traduction automatique
  - Pour les systèmes de Q/R
- Autres formalisations
  - L'*argumentative zoning*
  - La génération automatique de *slides*

Rappel sur la RST

- *Rhetorical Structure Theory*
- Modélisation de la structure rhétorique des textes
- Organisation hiérarchique de clauses
  - Relations discursives :
  - noyau vs. satellite : *arrière-plan*, *élaboration*, *préparation*...
  - multinucléaire : *contraste*

Rappel sur la RST





Applications basées sur le modèle de la RST

- En résumé automatique
  - Marcu, 1997, *From discourse structures to text summaries*
    - *rhetorical summarizer*
  - S'appuie sur la structure discursive des textes (division en noyau vs. satellite) pour décider de l'importance des phrases.
  - Plus les phrases sont proches de la racine, plus elles sont aptes à constituer un bon résumé du texte (niveau de l'arbre selon le niveau de granularité considéré).

Applications basées sur le modèle de la RST

- En recherche d'information
  - Corston-Oliver & Dolan, 1999, *Less is more: Eliminating index terms from subordinate clauses*
  - Optimisation des index : le document est seulement indexé sur les termes qui apparaissent dans les parties considérées comme les plus importantes.
  - L'information contenue dans les noyaux est considérée comme plus représentative du contenu du document que celle contenue dans les satellites :
    - on ne retient que les termes présents dans les noyaux
    - élimination des infos contenus dans les subordonnées

Applications basées sur le modèle de la RST

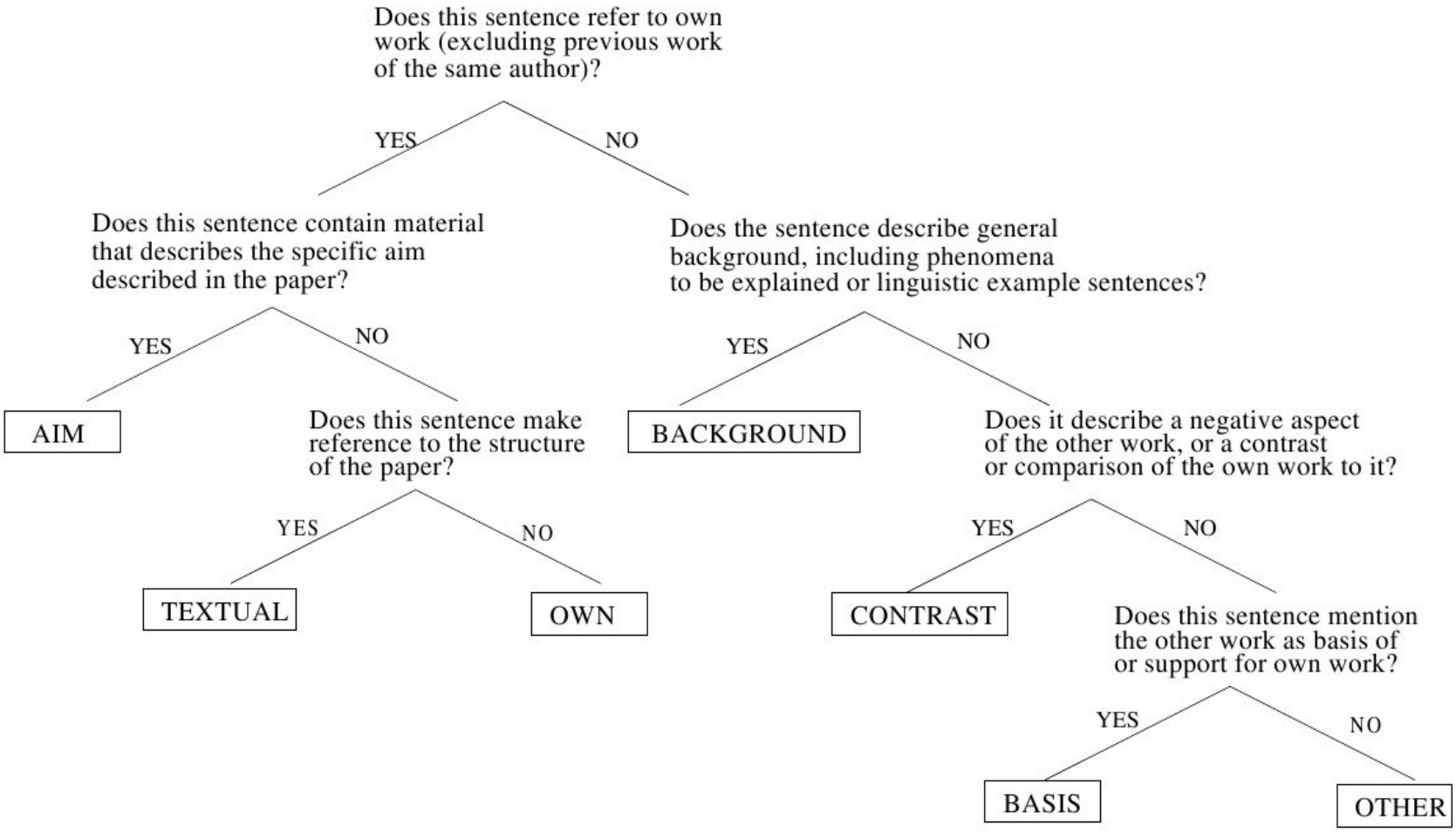
- En traduction automatique
  - Marcu & Carlson, 2000, *The automatic translation of discourse structures*
  - Aller au-delà de la traduction *phrase à phrase.*
  - Différente organisation de l'information selon les langues.
  - Traduction humaine → réorganisation des phrases pour satisfaire les contraintes de la langue-cible
  - Restituer le sens en s'adaptant aux modèles discursifs de la langue-cible
  - Ex : sur 115 relations CONTRAST en japonais → 34 CONTRAST, 27 ANTITHESIS/CONCESSION, 14 COMPARISON, 5 LIST... dans la traduction anglaise

Applications basées sur le modèle de la RST

- Pour les systèmes de Q/R
  - Verberne et al., 2006, *Discourse-based answering of why-questions*
  - Délimitation du segment de texte pouvant constituer la réponse
  - Certains types de relations sont plus à même de constituer des réponses à des *why-questions* : *cause, purpose, interpretation...*

Autres formalisations

- L'*argumentative zoning*
  - Teufel, Carletta & Moens, 1999, *An annotation scheme for discourse-level argumentation in research articles*
  - Détection des éléments structurels dans les articles scientifiques
  - Limites du repérage par cue-phrases :
  - "Unfortunately, this work does not solve problem X"
  - Problèmes liés à la considération de la phrase hors contexte
  - Typologie des relations rhétoriques adaptée au texte scientifique + construction de citation maps


Does this sentence refer to own work (excluding previous work of the same author)?
YES
NO
Does this sentence contain material that describes the specific aim described in the paper?
YES
NO
AIM
Does this sentence make reference to the structure of the paper?
YES
NO
TEXTUAL
OWN
Does the sentence describe general background, including phenomena to be explained or linguistic example sentences?
YES
NO
BACKGROUND
Does it describe a negative aspect of the other work, or a contrast or comparison of the own work to it?
YES
NO
CONTRAST
Does this sentence mention the other work as basis of or support for own work?
YES
NO
BASIS
OTHER

Autres formalisations

- La génération automatique de *slides*
  - Shibata & Kurohashi, 2005, *Automatic slide generation based on discourse structure analysis*
  - Sépare la phrase en *topic part* et *non-topic part*.
  - Réduit les *non-topic parts*.
  - Génère la diapositive.

(Due to the interruption of the three train services, JR Kobe-line, Hankyu Express Kobe-line and Hanshin Electric Railway, which connected between Osaka and Kobe, 450,000 people per day, 120,000 people per hour at the peak of rush, had no transportation. At the interruption sections in West Japan Railway Toukaidou Line, Sannyou Line, Hankyu Takarazuka, Imazu and Itami Line and Kobe-Electric Arima-line, transportation by alternate-bus was provided just after the earthquake occurred. From January 23th, when National Route 2 was opened, transportation by alternate-bus between Osaka and Kobe was provided. From January 28th, the alternate-bus priority lane was set up and smooth transportation was maintained.)

Railway Recovery (1)

- Interruption of the three train services, JR Kobe-line, Hankyu Express Kobe-line and Hanshin Electric Railway
  - 450,000 people per day, 120,000 people per hour at the peak of rush, had no transportation
- Interruption sections in West Japan Railway Toukaidou Line, Sannyou Line, Hankyu Takarazuka, Imazu and Itami Line and Kobe-Electric Arima-line
  - after the earthquake occurred
    - transportation by alternate-bus was provided
  - from January 23th, when National Route 2 was opened
    - transportation by alternate-bus between Osaka and Kobe was provided
  - from January 28th
    - the alternate-bus priority lane was set up and smooth transportation was maintained.

1. Introduction
   1. La relation de contraste
   2. TAL et Discours
      1. Enjeux, Méthodes et Limites
      2. Principales Application
2. **Repérage automatique de contraste**
   1. **Approche par marqueurs**
   2. Approche par lexicale
   3. Approche par parallélisme syntaxique
3. Conclusion et perspectives

Niveau d'analyse

- L'approche par marqueur implique un niveau d'analyse de type inter-propositionnel ou inter-phrastique.
  - Analyses proposées par la RST et la SDRT
- En TAL, ce niveau d'analyse est largement privilégié:
  - Analyse surfacique

Quelques travaux…

- Sporleder&Lascarides (2006)
  - Utilisation d'apprentissage automatique pour obtenir une classification à partir de données annotés manuellement
  - Corpus arboré :
    - BNC, North American News Text Corpus, English Gigaword Corpus
  - Extraction basée sur deux critères :
    - Identification de segments basée sur la présence de marqueurs discursifs
    - Détermination des frontières

Quelques travaux…

- Marcu (2000)
  - Repérage de ‘cue phrases ’ selon critères
    - Position
    - Cooccurrences avec indices typographiques
  - Annotation manuelle afin de caractériser les ‘cue phrases’ et leur capacité à introduire des relations de discours
  - Les résultats montrent que les connecteurs sont des indicateurs ambigus sur la structure rhétorique.

Repérage de relation de contraste dans des articles de SHS

- Corpus:
  - Revues scientifiques de SHS (projet Rhecitas)
  - Annoté morpho-syntaxiquement (Cordial)
  - Utilisation de la plateforme GATE
- Une ressource de marqueurs en français
- Expérimentation en corpus
- Evaluation des marqueurs discursifs

Corpus Rhecitas

- 247 articles issus du portail *revues.org*
  - AILE
  - Champ Pénal
  - Cybergeo
  - LIDIL
  - Terrain
- Documents au format XHTML
- Balisage et annotation des citations de référence
- Utilisation partielle de ce corpus …

Les marqueurs discursifs

- Deux sortes de marqueurs utilisés :
  - Les marqueurs argumentatifs (Schnedecker, 1998)
    - 5 triplets
    - *D'une part... d'autre part... d'autre part, d'abord ... ensuite...enfin*
  - Les connecteurs
    - 25 connecteurs
    - *Mais, bien que, cependant...*
- Ressource construite pour l'expérience ...

Expérimentation en corpus

- Construction de grammaires Jape (GATE) contenant les listes de connecteurs :

  - MACRO:C1

```
(
({Token.lemma=="or"})|
({Token.lemma=="certes"})|
({Token.lemma=="cependant"})|
({Token.lemma=="mais"})|
({Token.lemma=="pourtant"})|
({Token.lemma=="toutefois"})|
({Token.lemma=="néanmoins"})|
({Token.lemma=="malgré"})|
({Token.lemma=="contrairement"})
...
```

Expérimentation en corpus

- Projection des marqueurs discursifs sur le corpus
  - Première évaluation manuelle et ajout de nouveaux connecteurs
  - Deuxième ajout de ‘potentiels’ marqueurs discursifs se trouvant dans une phrase contenant des antonymes (les plus fréquents)
- Marquage de zones contenant un ou plusieurs marqueurs discursifs
  - Distinction dans le marquage de la position du connecteur dans la phrase
    - Début de phrase ou dans la phrase

Expérimentation en corpus

- Filtrage des zones ‘potentiellement’ contrastives contenant des citations

*Un premier saut théorique en ce sens sera la déterminante contribution de Spector et Kitsuse (1977) 31 , **bien qu**'il faille aussi mentionner les premiers travaux de Stanley Cohen (1973) , dans lesquels il s'intéresse au rôle des médias dans la création des Folk Devils and Moral Panics. [528_CHP]*

Evaluation

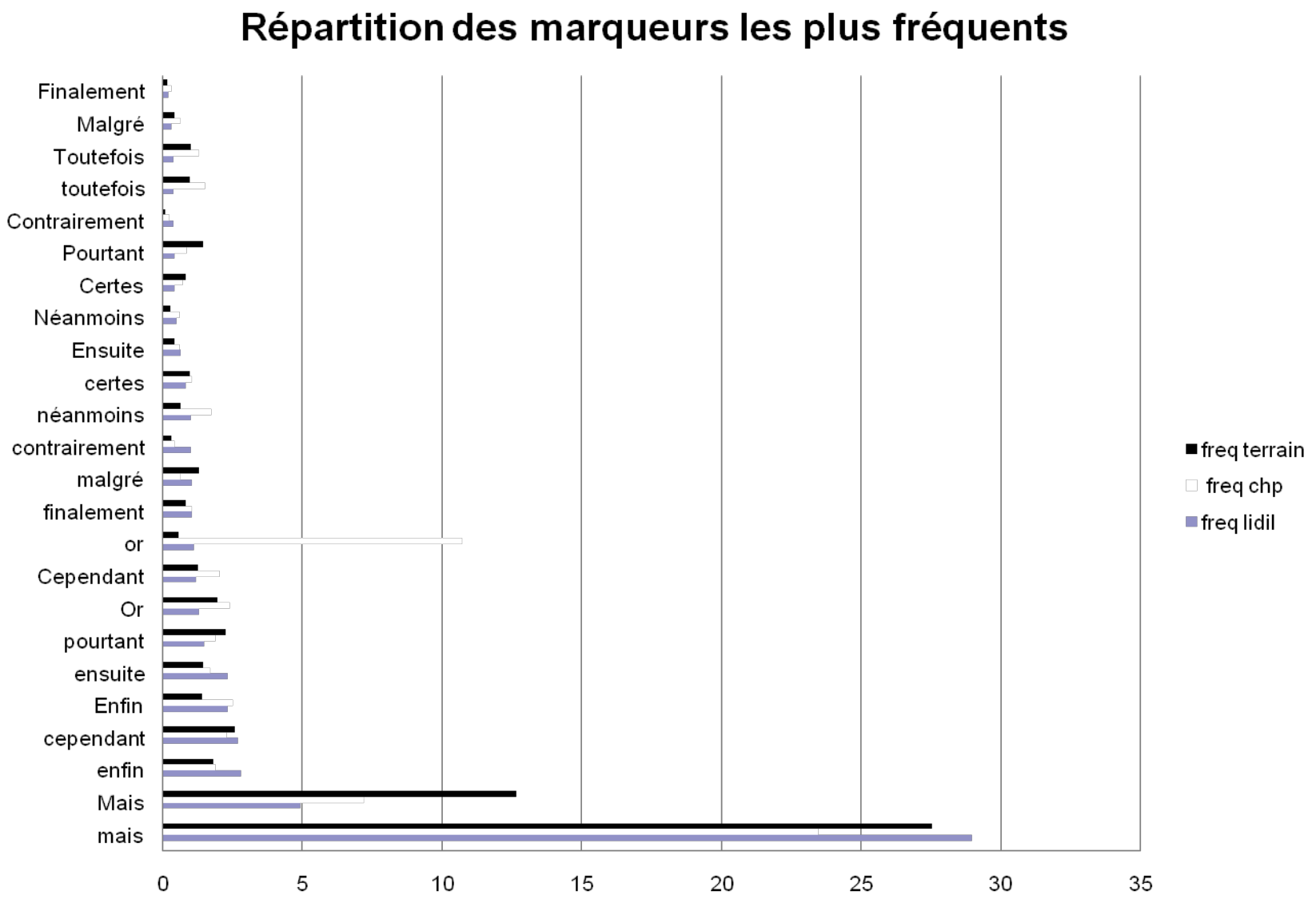
- Un très grand nombre de zones repérées, mais l'évaluation ne peut être effectuée
  - Pas de corpus annotés disponibles pour l'instant
- Mais possibilité d'évaluer les marqueurs
- Des connecteurs en grande quantité, une liste importante :
  - Sont-ils présents dans tous les corpus et en quelle proportion?
  - Quelle couverture du corpus pour les principaux connecteurs?

Evaluation

- Nombres de marqueurs discursifs par corpus :

Champ pénal	Terrain	lidil	TOTAL
3846	7 483	1 726	13 055

- Proportionnel à la grandeur de chaque sous-corpus
- Retour aux données et quantification des fréquences de chaque marqueur discursif.


Répartition des marqueurs les plus fréquents
Finalement
Malgré
Toutefois
toutefois
Contrairement
Pourtant
Certes
Néanmoins
Ensuite
certes
néanmoins
contrairement
malgré
finalement
or
Cependant
Or
pourtant
ensuite
Enfin
cependant
enfin
Mais
mais
0
5
10
15
20
25
30
35
freq terrain
freq chp
freq lidil

Discussion

- Les 24 marqueurs les plus fréquents sont les même dans les trois sous-corpus :
  - Stabilité dans la fréquence
- Le marqueur discursif *mais* est fortement surreprésenté dans tous les sous-corpus.
  - Environ 30% dans chaque corpus
- D'un point de vue général, les marqueurs en initiale de phrase semblent avoir un comportement plus stable.
- Quelques scories…

Contraste et antonymie

- Le contraste n'est pas forcément accompagné de *cue phrases* (74% des cas selon Marcu & Echihabi 2002) :
  - *Everyone assured us the offices would be **open** on Saturday. They were **closed**.*
  - *Ils ont un sens **étroit**, relativement précis et sur lequel chacun s'accorde à peu près, et un sens **large** qui ne fait pas l'unanimité,*
- → antonymie
- Opposition antonymique = contraste ?

Spenader & Stulp, 2007

- Spenader & Stulp, 2007, *Antonymy and Contrast Relations*
- Reconnaissance des relations de contraste à partir du repérage de paires d'antonymes adjectivales

(+ cooccurrence avec *but*).

- Analyse limitée à la phrase.

- Corpus issu du BNC (≈218 000 phrases).
- WordNet
  - antonymes directs
  - " indirects

Spenader & Stulp, 2007

- Antonymes directs :

  *Since 1973, in Columbus, Georgia, a death sentence has been sought for 43.8 per cent of those accused of killing a **white** female, and only 2.6 per cent of those accused of killing a **black** female.*

- Antonymes indirects :

  *Early work is often **missing** from an artist's oeuvre, while student work or juvenilia may be **saved** only by chance or possibly by a devoted family.*

Spenader & Stulp, 2007

- Antonymie non contrastive :

  *It has survived many more **recent** attempts by central government to have it replaced but since a major overhaul took place in 1986 this fine landmark has a secure **future**.*

- Ambiguïtés : *twisting/still* (ADJ ou ADV ?)

Spenader & Stulp, 2007

- **Antonymes sans *but***

Contrastive	+ Sense (32)	Source (23)	Direct	19
			Indirect	4
		Not Source (9)	Direct	4
			Indirect	5
	Wrong Sense (10)		Direct	0
			Indirect	10
			Total	45 (16%)
Not Contrastive	+ Sense (126)		Direct	103
			Indirect	23
	Wrong Sense (117)		Direct	30
			Indirect	87
			Total	243 (84%)
Full Total				288 Examples

Spenader & Stulp, 2007

- **Antonymie + *but***
- Directe :

  *He had **few** friends but **many** acquaintances.*

- Indirecte :

  *Eastern parts of England will start **bright** and mainly dry but central areas will be **cloudy** with showers in places.*

	All Antonyms	Direct	Indirect
Antonym Source of Contrast	177 (52%)	120 (68%)	57 (32%)
Antonym not Source of Contrast	41 (12%)	21 (51%)	20 (49%)
Wrong Sense	124 (36%)	38 (31%)	86 (69%)
Total	342 (100%)	179	163

Spenader & Stulp, 2007

- **Conclusion**
- Sur 14 110 phrases contenant *but* : 218 contenant une paire d'antonymes directs dont 177 sont à l'origine du contraste.

  → l'antonymie ne sert que rarement à marquer le contraste (-rappel)

  → combinés à *but*, une paire d'antonymes directs constituent une source de contraste dans 81% des cas (+précision)

L'antonymie comme source de contraste entre citations

- Repérer et analyser les cas où les antonymes signalent un contraste au sens de S. Teufel :

CONTRAST	Sentences contrasting own work to other work; sentences pointing out weaknesses in other research; sentences stating that the research task of the current paper has never been done before; direct comparisons

- Antonymes extraits du TLFi (7799 couples)
- Ensemble de marqueurs : néanmoins, cependant, pourtant, etc.
- Corpus d'articles scientifiques :
  - 265 textes (principalement criminologie et ethnologie)
  - 69 019 phrases
  - 2 246 372 mots

L'antonymie comme source de contraste entre citations

1 - Projection des antonymes :

- sur les phrases isolées (3,4% du total vs. 5% chez S&S)
- sur les couples de phrases adjacentes

2 - Filtrage avec les marqueurs et la présence de

	Phrases seules	Couples
Antonymes	2352	2195
Antonymes + 2 citations -	22	30
Antonymes + 2 citations + marqueurs	15	28

L'antonymie comme source de contraste entre citations

- Phrases seules :
  - Antonymes + 2 citations sans marqueur

    *En suivant les premiers ufologues, on aboutit, au **minimum** (Michel) à la constitution d'une nouvelle catégorie d'experts, d'une nouvelle classe de phénomènes ; au **maximum** (Guieu), à une retraduction complète de la société, à une nouvelle répartition des positions de porte-parole, d'experts.*

  → minimum/maximum

L'antonymie comme source de contraste entre citations

- Antonymes + 2 citations + marqueurs

  En attendant de trouver une autre manière de généraliser à partir de « contextes » par définition ***changeants*** *(Bensa 1996 : 44), c'est bien sous forme de systèmes non pas intemporels, mais simplement* ***stables*** *dans la moyenne durée que s'analysent les diverses versions anga de ce complexe associant l'individualisme, la domination masculine et l'ethos guerrier que Godelier (1982) a qualifié de logique du « Grand Homme »*

→ changeant/stable

L'antonymie comme source de contraste entre citations

- Couples
  - Antonymes + 2 citations sans marqueur
  - *Les lectes **élémentaires** (JANKA) ont servi pour la recherche des précurseurs. Dans les lectes **avancés** (FRANCA et MARIAN) l'intégration des particules dans une syntaxe complexe a pu être retrouvée.*

  → élémentaire/avancé

L'antonymie comme source de contraste entre citations

- Antonymes + 2 citations + marqueurs

En revanche, si la folie désassimile et aliène, si elle rend le moi inaccessible à soi et à la société, le fou peut être déclaré ***irresponsable*** *et ne peut être poursuivi pénalement (Tarde, 1890). Le pervers, en revanche, est* ***responsable*** *: le pervers, le fou moral, n'est nullement un aliéné (Tarde, 1890).*

→ irresponsable/responsable

L'antonymie comme source de contraste entre citations

- Centralité de la paire d'antonymes dans la relation de contraste

	Phrases seules			**Couples**		
Antonymes + 2 citations - marqueurs	occ	contraste	anto	occ	contraste	anto
	22	13 (59%)	11 (84%)	30	15 (50%)	10 (67%)
Antonymes + 2 citations + marqueurs	occ	contraste	anto	occ	contraste	anto
	15	8 (53%)	5 (62.5)	28	18 (64%)	12 (67%)

L'antonymie comme source de contraste entre citations

- Résultats peu significatifs :
  - Difficile de calculer l’influence des marqueurs
  - Peu de résultats
  - Dictionnaire d’antonymes mal adapté

« Pour encoder les relations locatives , les enfants *nominaux* utilisaient *des noms* pour indiquer le lieu (papa bureau) , **tandis que** les enfants *pronominaux* utilisaient *les pro formes* déictiques (here et there). »

- connecteur « `tandis que` »
- termes contraires « `nominal` » vs « `pronominal` »
- parallélisme syntaxique: `les enfants ADJ utilisaient DET N`

"Our intuition is that similarities concerning the surface, the content and the structure of textual units can be a way for authors to explicit their intention to consider these units with the same rhetorical importance"

(Guégan, Hernandez, 2006)

Parallélisme et contraste:

- Cadres de discours (Charolles, 1997)
- Enumérations (Luc et al. 1999)
- Titres (Summers, 1998)

Prise en compte des parallélismes syntaxiques pour le repérage automatique de relations du discours (1):

- Génération automatique de transparents (Shibata et Kurohashi, 2006)
  - Similarité entre mots basée sur
    - Identité de forme
    - Identité de lemme
    - Identité de catégorie syntaxique
    - Distance dans un thesaurus
  - Une Clause/Phrase est une chaine de mots
    - Similarité entre chaines
    - Choix de relation (liste/contraste) basée sur la similarité des *topics*
  - C'est du japonais:
    - Marquage casuel → analyse syntaxique automatique performante
    - Marquage du topic (–wa) → identification du topic
    - Tête-à-droite → segmentation facilitée (prédicat = frontière)

Prise en compte des parallélismes syntaxiques pour le repérage automatique de relations du discours (2):

- Contraste et Parallélisme (Widlocher, 2008)
  - Similarité entre phrases base sur:
    - Bigrammes de mots pleins (N,A,V)
    - Identité de lemmes
    - Identité de position dans la phrase

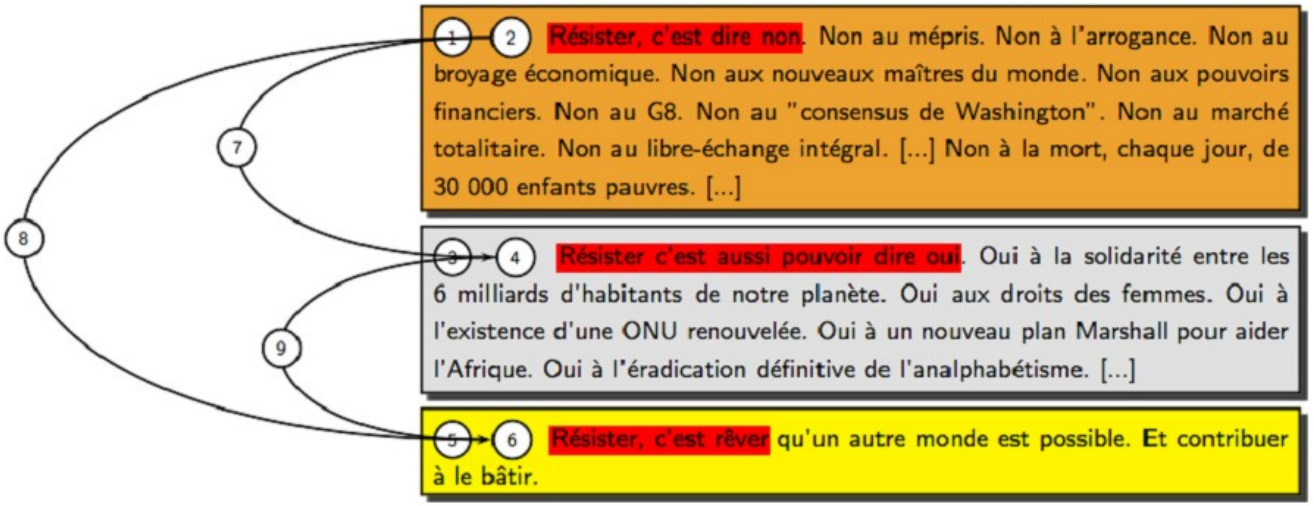


FIG. 9.13 – Contraste et parallélisme structurel

Parallélismes multi niveaux, pour repérer des relations coordonnantes:

- Projet ANNODIS
- Analyse syntaxique automatique par SYNTEX (Bourigault, 2007)

Principe générale:

- Unité: phrase
- Construction d'un ensemble de triplets suivant les liens de dépendance
  - `X - Gouv(X) - Gouv(Gouv(X)),`
  - `X - Gouv(X) - Dep(Gouv(X))`
- Niveau d'abstraction variable
  - `LEM-LEM-LEM, LEM-LEM-CAT,...,LEM-CAT-CAT, CAT-CAT-CAT`
- Calcul du taux de recouvrement de triplets entre phrases

Illustration du fonctionnement:

"assassin"	"héros"
P0:assassin#être#vous	P0:héros#être#vous
P1:assassin#être#Pro	P1:héros#être#Pro
P1:assassin#VCONJS#vous	P1:héros#VCONJS#vous
P1:Nom#être#vous	P1:Nom#être#vous
P2:assassin#VCONJS#Pro	P2:héros#VCONJS#Pro
P2:Nom#VCONJS#vous	P2:Nom#VCONJS#vous
P2:Nom#être#Pro	P2:Nom#être#Pro
P3:Nom#VCONJS#Pro	P3:Nom#VCONJS#Pro

Recouvrement
Dans l'exemple: 50%
Des deux phrases: 45,4%

Expériences sur corpus:

- recherche de paires de phrases similaires se situant dans le même paragraphe
  - Contraste entre phrases
    - CONTRASTE(ELABORATION(**1**,***2***),ELABORATON(**3**,***4***))

Le concept d'innovation est moins exigeant que celui d'évolution**.** ***Il sert surtout à indiquer ce qui est déviant par rapport à une structure (centrale) prise comme point de référence initial.*** Le concept d'évolution, lui, est beaucoup plus global que celui d'innovation**.** ***Il sert à observer ce qui se passe au plan des changements des structures centrales proprement dites***.

(Champ Pénal)

Expériences sur corpus:

- recherche de paires de phrases similaires se situant dans le même paragraphe
  - Listes

Que faire de ces brutes, sortes d'animaux humains en cage ? Les amender ? Comment ? **Par le pli de la bonne habitude (travail, silence?) ? Par l'intimidation ? Par l'exhortation morale ? Par le changement total de vie et de milieu (transportation?) ? Par l'isolement (régime cellulaire) ?**

(Champ Pénal)

Expériences sur corpus:

- Similarité entre titres:

  - Point de vue biologique
  - Point de vue comportemental
  - Point de vue zootechnique

  - Le bloc de l' Ouest
  - Le bloc de l' Est

  - Les arguments contre ces bombardements
  - Les arguments pour ces bombardements

  - Hiroshima durant la Seconde Guerre mondiale
  - Nagasaki durant la Seconde Guerre mondiale

(Wikipedia)

Problèmes de l'approche par taux de recouvrement de triplets et solutions envisagées:

- Problèmes:
  - Trop influencé par des répétitions lexicales
  - Trop dépendant de la segmentation en amont
  - La phrase comme unité ne convient pas
  - *Résultats difficilement exploitables*

- Solutions:
  - Définir les unités liées par une relation de parallélisme comme les séquences maximales ressemblantes.
    - Approche par n-grammes
    - S'affranchir de la phrase comme unité
  - Intégrer l'architecture textuelle dans le calcul
    - Classer les parallélismes en fonction de la similarité de leurs positions dans le texte

1. Introduction
   1. La relation de contraste
   2. TAL et Discours
      1. Enjeux, Méthodes et Limites
      2. Principales Applications
2. Repérage automatique de contraste
   1. Approche par marqueurs
   2. **Approche par antonymie**
   3. **Approche par parallélisme syntaxique**
3. Conclusion et perspectives

Bilan

- Relation de contraste
- TAL et discours :
  - Principaux enjeux
  - Applications
- Expérimentation de différentes approches
  - Repérage par marqueurs
  - Repérage par paires antonymes
  - Repérage par parallélisme syntaxique

Perspectives pour l'analyse automatique du discours

- Maturation des théories du discours et meilleure adaptation pour le TAL
- Méthodes d'analyses hybrides :
  - Prise en compte de l'ensemble d'observables à notre portée (approches multi-niveaux)
  - Création de ressources
- Mieux définir les besoins applicatifs

Références

- **Bourigault, D.,** (2007), *Un analyseur syntaxique opérationnel : SYNTEX*. Mémoire d'habilitation à diriger les recherches. Université Toulouse-Le Mirail.
- **Busquets, J.,** (2007) *Discourse Contrast : Types an Tokens* , In Language, Representation and Reasoning. Memorial Volume to Isabel Gómez Txurruka, M. Aurnague. K. Korta and J. M. Larrazabal (eds.). Bilbao : University of Basque Country Press, Charolles, M., (1997), L'encadrement du dicours - Univers, champs, domaines et espace, Cahier de recherche linguistique, 6, pp. 1-60.
- **Corston-Oliver & Dolan,** (1999), *Less is more: Eliminating index terms from subordinate clauses*
- **Guégan, M. et Hernandez N.,**(2006), *Recognizing Textual Parallelisms with Edit Distance and Similarity Degree*, EACL 2006, Trento, Italie
- **Lakoff, R.,** (1971), *If 's, and's, and but's about Conjunction*. in Fillmore, C., and Langendoen, D. (eds.), Studies in Linguistics Semantics, Holt, Reinhart and Wilson, New York, 115-150.
- **Marcu, D.,** (1997), *From discourse structures to text summaries*
- **Marcu, D.,**(2000), *The Rhetorical Parsing of Unrestricted Texts: A Surface-Based Approach*.Computational Linguistics,26,395-448.
- **Marcu & Carlson,** (2000), *The automatic translation of discourse structures*
- **Schnedecker,C.,** (1998), *Les corrélats anaphoriques. Recherches Linguistiques*.
- **Shibata, T. and Kurohashi, T.,** (2005),*Automatic Slide Generation Based on Discourse Structure Analysis.* In Proceedings of Second International Joint Conference on Natural Language Processing (IJCNLP-05),
- **Verberne et al.,** (2006), *Discourse-based answering of why-questions*
- **Sporleder, C. and Lascarides, A.,** (2006). *Using Automatically Labelled Examples to Classify Rhetorical Relations: An Assessment*, to appear in Natural Language Engineering.
- **Teufel, Carletta & Moens,** (1999), *An annotation scheme for discourse-level argumentation in research articles*
- **Widlöcher, A.,** (2008) *Analyse macro-sémantique des structures rhétoriques du discours - Cadre théorique et modèle opératoire* Thèse de doctorat, Caen.