

Alternating Modifiers in Nominal Paraphrases: Investigation in Specialised/Lay Comparable Corpora

Bruno Cartoni (Limsi-CNRS(Fr), University of Geneva, (CH))

Louise Deléger (Limsi-CNRS (Fr))

Headline

- ▶ **Objective: Exploring alternating nominal paraphrases between lay and specialised discourses**
 - ▶ Confirm linguistic hypothesis
 - ▶ Fine-tune paraphrase pattern
- ▶ **Description of alternating paraphrase**
- ▶ **Methods of investigation**
 - ▶ Monolingual comparable corpora
 - ▶ Lexical-based pattern of alternating paraphrase
 - ▶ Computing indexes to compare preferences
- ▶ **Results and analysis**
- ▶ **Conclusion**

Objectives

- ▶ Nominal paraphrases in lay & specialised discourse
- ▶ Paraphrase with the same nominal head, but different (and morphologically related) modifiers.
- ▶ 3 different N + modifiers studied:
 - ▶ N + Relational adjectives
 - ▶ N + Prefixed relational adjectives
 - ▶ N + Deverbal adjectives
- ▶ The three can be paraphrased by “N + prep + N”

Relational adjectives

- ▶ In French, relational adjectives are typical of specialised discourses (L'homme 2004, Daille 1999)
 - ▶ Specialised: *infarctus myocardique* [EN: myocardial infarction]
 - ▶ Lay: *infarctus du myocarde* [EN: infarction of the myocardium]
- ▶ Relational adjectives:
 - ▶ Derived from nouns
 - ▶ Suffixed in *-al, -ien, -ique, ...*
 - ▶ Denote the relation between the noun they modify and their base noun

muscle abdominal [EN: abdominal muscle]

→ *muscle of the abdomen*

Prefixed relational adjectives

- ▶ When prefixed, relational adjectives are the formal base of the adjectives, but on the semantic side, the prefixation rule applies to the nominal base (Fradin, 2008).

anticancéreux [EN: anticancerous]

- ▶ prefixation in “anti” actually applies to the base noun *cancer*
 - ▶ *Un traitement anticancéreux* [EN: an anticancerous treatment]
 - ▶ *Un traitement contre le cancer* [EN: a treatment against cancer]
- ▶ Prefix and preposition are (not always) semantically related

Deverbal adjectives

- ▶ Deverbal adjectives: suffixed in *-if*, *-eur* and *-oire*
- ▶ Alternatively used in place of a deverbal noun that shares the same verbal base

troubles dépressifs [EN: depressive disorders]

troubles de la dépression [EN: disorders of depression]

Methods of investigation: Monolingual comparable corpora

- ▶ **Monolingual comparable corpora:**
 - ▶ Same genre, same topic, same languages
 - ▶ Different types of register for different target audience (Specialist Vs. Lay)
- ▶ **Medical corpora**
 - ▶ Topics: cancer, tobacco use, diabetes
 - ▶ Gathered from various sources (Deléger & Zweingenbaum 2009)
 - ▶ Tag and lemmatised
- ▶ **Stats:**

	Specialised	Lay
Sentences	65,990	99,028
Token	1,414,090	1,414,805

Methods of investigation:

Gathering pairs of adjective - noun

- ▶ All (prefixed) relational/deverbal adjectives extracted from both sides of the corpus (using typical suffixes)
- ▶ Analysed by a French morphological analyser (Dérif)
- ▶ Manually checked

- ▶ **3 lists of pairs**

Rel_adj - Nom

prostatique – prostate
abdominal – abdomen
tumoral – tumeur
...

873 pairs

Pref_rel_adj - Nom

antigrippe – grippe
intraveineux – veine
...

303 pairs

deverb_adj - Nom

depressif – depression
opérateur – opération
...

214 pairs

Methods of investigation:

Lexical-based pattern of alternating paraphrase

- ▶ For each pair of the lists, we look into both sides of the corpus for equivalent NPs (i.e with the same headnoun)

$$N_2 + \text{Adj}_1 \leftrightarrow N_2 + \text{prep} + (\text{det}) + N_1$$

where Adj_1 and N_1 are from the pairs lists

- ▶ Eg. : *prostatique* – *prostate*
 - ▶ *tumeur prostatique* – *tumeur de la prostate*
 - ▶ *cancer prostatique* – *cancer de la prostate*

Methods of investigation:
Computing indexes to compare preferences

$$I = \frac{\text{sum (N+Adj)}}{\text{sum (N+Adj) + sum (N+prep+N)}}$$

- ▶ Following our hypothesis, the preference for N +Adj should be higher (close to 1) in the specialised side of the corpus than in the lay side (close to 0)

Results : quantitative analysis

▶ Reminder:

Rel_adj - Nom	873 pairs
Pref_rel_adj - Nom	303 pairs
Deverb_adj - Nom	214 pairs

▶ Mean preference index

	Relational (547 pairs)	Prefixed (16 pairs)	Deverbal (36 pairs)
Spec.	0.73	0.80	0.74
Lay	0.43	0.24	0.33

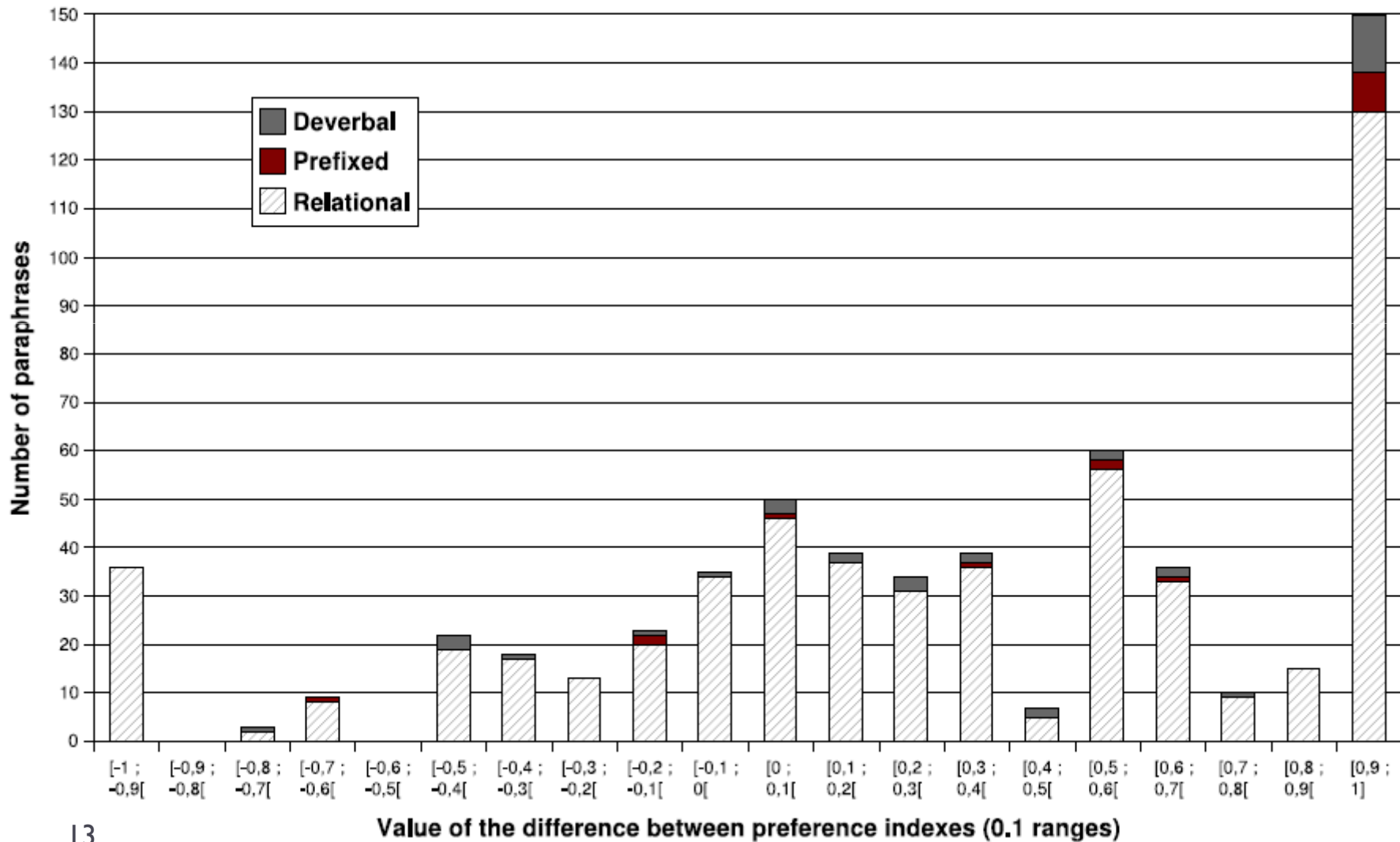
▶ The hypothesis is confirmed, but ...

Results: quantitative analysis

	Index		Difference
	S	L	
(1) volume tumoral / volume de la tumeur (<i>tumoral volume / volume of the tumor</i>)	1	0	1
(2) risque infectieux / risque d'infection (<i>infectious risk / risk of infection</i>)	0.89	0.06	0.83
(3) surveillance glycémique / surveillance de la glycémie (<i>glycemic control / control of glycemia</i>)	1	0.4	0.6
(4) injection intraveineuse / injection dans la veine (<i>intravenous injection / injection in the vein</i>)	1	0	1
(5) vaccination antigrippale / vaccination contre la grippe (<i>anti-flu vaccine / vaccine against the flu</i>)	0.62	0	0.62
(6) mouvement respiratoire / mouvement de la respiration (<i>respiratory movement / movement of respiration</i>)	1	0	1
(7) hormonothérapie substitutive / hormonothérapie de substitution (<i>substitutive hormone therapy / hormone therapy of substitution</i>)	0.67	0	0.67

- ▶ Computing the differences between I(S) and I(L)
 - ▶ Classifying the differences
-

Results: Classified Differences



Result: focus on zero and negative values

▶ Zero differences

- ▶ N+Adj → some are very common in lay language
rythme cardiaque [EN: cardiac rhythm]

▶ Negative differences

- ▶ N+Adj is stronger in lay languages than in specialised one
- ▶ This happens when the *Noun phrase* does not belong to the specific specialised language (i.e. medicine)
 - ▶ campagne publicitaire / campagne de publicité
[EN: advertising campaign]
 - ▶ besoin communautaire / besoin de la communauté
[EN: community need]

Result: a focus on preposition

- ▶ The relation between Relational Adjective and Head Noun is said to be semantically “undetermined”
 - ▶ Preposition mostly used: *de (of)*
- ▶ Sometimes, 1 N+Adj \leftrightarrow many N+prep+N
 - ▶ *dépendance alcoolique* \leftrightarrow *dépendance à l'alcool, dépendance envers l'alcool*
- ▶ Preposition meaning related to Prefix meaning (but not always)
 - ▶ *lutte antitabac* \leftrightarrow *lutte contre le tabac*
 - ▶ *législation antitabac* \leftrightarrow *législation sur le tabac*
 - ▶ *tumeur intraprostatique* \leftrightarrow *tumeur de la prostate*

Conclusion

- ▶ **Conclusion on the hypothesis**

- ▶ Noun phrases with adjectival modifiers are preferred by specialised language
- ▶ Mainly when they are part of the specialised vocabulary of the domain

- ▶ **Conclusion for the paraphrases**

- ▶ Useful pattern : distinction between registers
- ▶ Preposition is important in the pattern : it should not be restricted to the “de” (of) preposition
- ▶ Paraphrases of “prefixation” is not necessarily straightforward

- ▶ **Further work : give more generality to this study**

- ▶ Build larger corpora

-
- ▶ ¹⁶ Study other domains

Thank you

bruno.cartoni@unige.ch // louise.deleger@limsi.fr