

Adjectives in Treebank

Anna Kupść

`Anna.Kupsc@u-bordeaux3.fr`

UE TAL

November 28, 2008

Valence Lexicon (for Adjectives)

- valence, subcategorisation frame
- combinatory potential of a predicate: number and type of arguments
- often created for verbs but other categories can have syntactic arguments as well \rightsquigarrow adjectives
- predicate and language specific:

FR: fiers de ses enfants PP[de]

AN: proud of their children PP[of]

PL: dumni ze swoich dzieci PP[z]

Objective and Motivation

Objective: automatically create a valence lexicon for French adjectives for automatic text processing

Motivation:

- language learning and acquisition
- NLP: analysis, generation
- applications: MT, IE, ...
- not available for French

Method

- automatic extraction from a corpus
- corpus: treebank of Paris7, *Le Monde*, about 1 million words
- rich annotations: morphosyntactic, constituents and functions (for arguments of verbs)
- validated by human experts

Treebank

text + different type of information (annotations/tags)

- document structure: segmentation in paragraphs and sentences
- tokenization: simple and compound words
- morphological: category (POS), lemma, number, gender, person . . .
- syntactic: constituents
- functions: grammatical functions of major constituents (dependents of **verbs**)

XML tags

Paul aime Marie

annotations: flat structure

```
<SENT>
<NP fct="SUJ">
  <w cat="N" ee="N-P-ms" lemma="Paul">Paul</w></NP>
<VN>
  <w cat="V" ee="V--P3s" lemma="aimer">aime</w></VN>
<NP fct="OBJ">
  <w cat="N" ee="N-P-fs" lemma="Marie">Marie</w>
</NP>
</SENT>
```

Paul aime Marie

annotations: flat structure

```
<SENT>
<NP fct="SUJ">
  <w cat="N" ee="N-P-ms" lemma="Paul">Paul</w></NP>
<VN>
  <w cat="V" ee="V--P3s" lemma="aimer">aime</w></VN>
<NP fct="OBJ">
  <w cat="N" ee="N-P-fs" lemma="Marie">Marie</w>
</NP>
</SENT>
```

Paul aime Marie

annotations: flat structure

```
<SENT>  
<NP fct="SUJ">  
  <w cat="N" ee="N-P-ms" lemma="Paul">Paul</w></NP>  
<VN>  
  <w cat="V" ee="V--P3s" lemma="aimer">aime</w></VN>  
<NP fct="OBJ">  
  <w cat="N" ee="N-P-fs" lemma="Marie">Marie</w>  
</NP>  
</SENT>
```


Paul aime Marie

annotations: flat structure

```
<SENT>  
<NP fct="SUJ">  
  <w cat="N" ee="N-P-ms" lemma="Paul">Paul</w></NP>  
<VN>  
  <w cat="V" ee="V--P3s" lemma="aimer">aime</w></VN>  
<NP fct="OBJ">  
  <w cat="N" ee="N-P-fs" lemma="Marie">Marie</w>  
</NP>  
</SENT>
```

Types of Adjectives

four subtypes distinguished:

- qualitative: *original, intéressant, incroyable, facile*, etc.
- numerals: *trois, troisième, 3*
- quantifiers: *plusieurs*
- adjectival pronouns: *quel*

⇒ only qualitative adjectives can have arguments

Annotations

- (1) [*NP* la [*A* moindre] [*N* réforme]]
- (2) [*NP* les [*A* dix] [*A* derniers] [*N* mois]]
- (3) [*NP* la [*AP* [*Adv* plus] [*A* grande]] [*N* discrétion]]
- (4) [*NP* Cette [*N* comparaison] [*AP* [*A* préliminaire]]] semble
 SUJ
 [*AP* [*A* valable]]
 ATS
- (5) [*NP* le [*Adv* plus] [*A* froid] [*PP* des [*NP* [*A* dix] [*A* derniers]
 [*N* mois]]]]

Complements of Adjectives

three possible realisations in French:

- (6) sûr [*PP* de sa réussite]
- (7) sûr [*S_{Sub}* qu'il réussira]
- (8) sûr [*VP_{inf}* de réussir]

not always a complement, of course . . .

Comparative Constructions

- (9) La réunion était [_{AP} **plus** intéressante [_{Ssub} **que** [_S je ne pensais]]].
- (10) La réunion était [_{AP} **plus** intéressante [_{Ssub} **que** [_{NP} l'année dernière]]].

⇒ check adverbs

“Intensifier” Constructions

(11) Cette histoire est [_{AP} *(**trop**) fabuleuse [_{VP_{inf}} pour être vraie]].

other adverbs: *assez*, *bien*, *suffisamment*

(12) Antoine est [_{AP} *(si) inquiet [_{S_{sub}} qu’il n’ ose rien demander]].

other adverbs: *tellement*, *à ce point*

⇒ check adverbials

Extraposition

- (13) Paul est heureux [*Ssub* que Marie vienne].
- (14) C'est agréable [*Ssub* que Marie vienne].
⇒ [*Ssub* Que Marie vienne] est agréable.
- (15) Paul est capable [*VPinf* de sortir tous les jours].
- (16) C'est agréable [*VPinf* de sortir].
⇒ [*VPinf* (De) sortir] est agréable.

In the Corpus

- (17) Subject Extraposition:

[_V C'est] [_{AP} agréable] [_{Ssub} que Marie vienne].
SUJ ATS OBJ

- (18) Ssub complement:

[_{NP} Paul] [_V est] [_{AP} heureux [_{Ssub} que Marie vienne]].
SUJ ATS

- (19) ATO (object complement):

[_{NP} Jean] trouve [_{AP} triste] [_{Ssub} que Marie parte].
SUJ ATO OBJ

Tough Adjectives

- (20) [*NP* Ses erreurs]_i sont faciles [*VP_{inf}* à comprendre --_i].
- (21) Il est facile [*VP_{inf}* **de** comprendre ses erreurs].
- (22) [*VP_{inf}* (De) comprendre ses erreurs] est facile.
- (23) [*NP* Jean] est long [*VP_{inf}* à comprendre ses erreurs].

⇒ extracted phrases not annotated in the corpus

Superlative Constructions

(24) [_{NP} la [_{AP} plus sévère] [_N récession] [_{PP} parmi les
Douze]]

other prepositions: *entre*, *de*

⇒ PP is not part of AP

Positive Constructions

(25) une pierre [*AP* noire [*PP* comme l'enfer]]

⇒ PP[comme] shouldn't be considered the complement

PrepLex

- lexicon which specifies argumental and non-argumental prepositions
- 49 argumental prepositions: simple and complex
- **comme**: non-argumental for adjectives (comparatives)
- several complex prepositions not in PrepLex \Rightarrow excluded:
à la tête de, à la limite de, à la suite de, par l'intermédiaire de

Notation

- the subject (a semantic or a syntactic argument) vs. complements
- function and category
- the subject (SUJ): NP, VPinf, Ssub (SsubI or SsubS)
- complements: OBJ (SsubI or SsubS), P-OBJ (PP+prep or VPinf+complementizer), cl (non-mapped clitics)
- basic frame: SUJ:NP

Numbers

- 2153 qualitative adjectives (types), or 16410 occurrences (tokens)
- 86% of all types (1849 adjectives; 11116 occurrences) only with the basic frame
- 304 adjectives (5294 occurrences) with a different or an additional frame than basic:
 - 238 adjectives have an additional frame
 - 66 adjectives have a different frame (no basic)
- 41 frames in general (including the basic one)

Frames

FRAME	freq.	#adjs	SINGLETON FRAMES (freq. 1)
SUJ:NP (basic)	15485	2087	SUJ:VPinf[de] P-OBJ:PP[pour]
SUJ:NP P-OBJ:PP[à]	278	81	SUJ:NP P-OBJ:PP[devant]
SUJ:NP P-OBJ:PP[de]	204	94	SUJ:NP P-OBJ:PP[face à]
SUJ:NP P-OBJ:VPinf[de]	83	44	SUJ:NP P-OBJ:PP[sous]
SUJ:VPinf[de]	66	29	SUJ:Ssub[que]
SUJ:NP P-OBJ:VPinf[à]	53	16	SUJ:NP P-OBJ:PP[à] P-OBJ:VPinf[de]
SUJ:NP P-OBJ:PP[pour]	35	29	SUJ:NP P-OBJ:PP[jusqu'à]
SUJ:NP P-OBJ:PP[en]	30	23	SUJ:NP OBJ:VPinf
SUJ:NP P-OBJ:VPinf[pour]	24	6	SUJ:NP P-OBJ:PP[à] P-OBJ:VPinf[sans]
SUJ:NP P-OBJ:PP[dans]	22	14	SUJ:Ssub[que] P-OBJ:PP[en]
SUJ:Ssubl[que]	18	11	SUJ:NP OBJ:Sint
SUJ:NP OBJ:Ssub[que]	18	4	SUJ:NP P-OBJ:PP[selon]
SUJ:NP P-OBJ:PP[par]	13	12	SUJ:NP P-OBJ:PP[envers]
SUJ:NP OBJ:Ssubl[que]	12	3	SUJ:NP P-OBJ:PP[de] P-OBJ:PP[à]
SUJ:NP P-OBJ:PP[sur]	11	11	SUJ:NP P-OBJ:PP[vis-à-vis de]
SUJ:NP P-OBJ:PP[avec]	9	6	SUJ:NP P-OBJ:PP[parmi]
SUJ:NP P-OBJ:PP[loc]	8	8	SUJ:VPinf[à]
SUJ:NP P-OBJ:PP[entre]	5	3	SUJ:NP P-OBJ:PP[dans] cl:me/OBJ
SUJ:SsubS[que]	6	5	
SUJ:NP P-OBJ:PP[chez]	4	3	
SUJ:NP P-OBJ:PP[depuis]	3	3	
SUJ:VPinf[de] P-OBJ:PP[à]	3	3	
SUJ:NP P-OBJ:PP[après]	2	2	

Distribution of (Multi)frame Adjectives

# frames	# adjectives
9	1
6	4
5	4
4	14
3	34
2	186
1	1910

(*difficile*)

(*nécessaire, facile, indispensable, rare*)

(*présent, nombreux, élevé, possible*)

Basic Frame

- among 304 “interesting adjectives”, 238 appear with the basic frame
- the use of the basic frame is prevailing: only 24 adjectives use more often a different frame
- ratio (**basic/other realisations**)
 absurde (1/3), analogue (1/3), **capable** (7/25), comparable (3/7), compatible (2/5), **conforme** (1/9), conscient (1/13), constitutif (1/3), content (1/3), créancier (2/6), digne (1/3), dépourvu (1/4), **enclin** (1/3), inférieur (7/45), inscrit (1/3), insensible (1/3), **originnaire** (1/3), proche (11/38), prêt (5/30), soucieux (2/15), spécialiste (1/4), supérieur (20/61), âgé (4/10), égal (2/12)

More Than the Subject

- 66 adjectives never appear with the subject alone
- **examples**: accessoire, accompagné, adhérent, admis, agrégé, allergique, amateur, apte, aride, attendant, avare, concessionnaire, condamné, **coupable**, coutumier, destructeur, **distant**, **désireux**, **exempt**, **fier**, fixé, incapable, interdit, semblable
- an obligatory complement?
 - yes: *exempt*
 - no (insufficient data): *coupable, distant, fier, désireux* ...

VPinf[pour]

6 adjectives:

- correct: *indispensable, isuffisant, suffisant, nécessaire*
- wrong: *énorme, étroit*

SUJ:NP|P-OBJ:PP[loc]

8 adjectives:

- *y* with P-OBJ function
- *facile, fort, hermétique, nombreux, rare, réduit, technique, élevé*
- not all of them really take a P-OBJ complement . . .

Two Complements

4 adjectives:

- *heureux*: SUJ:NP|P-**OBJ:PP[à]**|P-OBJ:VPinf[de]
- *créateur*: SUJ:NP|P-**OBJ:PP[à]**|P-OBJ:VPinf[sans]
- *solidaire*: SUJ:NP|P-**OBJ:PP[dans]**|**cl:me/OBJ**
- *supérieur*: SUJ:NP|P-**OBJ:PP[de]**|P-OBJ:PP[à]

Propositional Arguments

- semantico-syntactic classification of adjectives (based on argument types):
 - Ssubl and/or VPinf
 - SsubS and/or VPinf
 - VPinf only

(28) **SUJ:VPinf[de]**: impossible, impératif, **indispensable**, interdit, intéressant, inutile, nécessaire, ...

(29) **SUJ:Ssubl[que]**: acquis, certain, clair, **fréquent**, indéniable, inévitable, probable, surprenant, sûr, urgent, vrai

(30) **SUJ:SsubS[que]**: compréhensible, **fréquent**, **indispensable**, logique

fréquent+SUJ:SsubS with *peu* (a semi-negative adverb)

SUJ:NP|OBJ:Ssub[que]

4 adjectives:

- *autre, tel*: resemble comparatives but their form is positive
- *meilleur, pire*: irregular comparative forms

Ssub indicates any phrase introduced by *que*, not necessarily a true subordinate clause

Summary of Results

- an extraction method based on linguistic knowledge and corpus annotations + PrepLex
- quality of results (no quantitative evaluation):
 - productive constructions vs. adjective valence
⇒ VPinf and Ssub arguments
ok + comply with the semantico-syntactic classification of adjectives
 - PP complements:
 - PrepLex ⇒ insufficient: every argumental preposition can also introduce a non-argument (context)
 - conversion of clitic arguments ⇒ imperfect: corpus annotations

Future

- PP complements: more data + statistical methods (correlations)
- VPinf:
 - explore corpus annotations, ex. subjectless phrases: (il pourrait être) *Impossible d'ignorer les liaisons transatlantiques*
 - identify *tough* adjectives by using a valence lexicon for verbs
- quantitative evaluation

available on-line:

http://erssab.u-bordeaux3.fr/article.php3?id_article=150