Nouns governing *that*-complement clauses : exploring the syntax / semantics interface

Nicolas Ballier USPC Paris Diderot CLILLAC-ARP (EA3967)

2/2/2018

OUTLINE

- An almost plausible agreg question
- the *modus*/*dictum* analysis proposed in (Ballier 2012) and developed in (Kanté 2013) and (Kanté 2016).
- Some reminders about lexical databases
- The wordnet Lexical DataBase
- Some Models (Valency Grammar: Herbst et al. 2004)
- Some more questions

PS: check the FRAMENET Resource

This presentation is aimed at presenting an endearing obsession with nouns governing complement clauses (sometimes called ‘complement taking-nouns’)

A companion handout sums up some of the corpus findings that may contradict some of the commonly held assumptions about those noun complement clauses

AIMS

- question the semantic singularity of nouns that can be used with *that*-complement clauses, of the type *the fact that* (but also *hypothesis*, *riddle*, *clue*)
- qestion the "derivational hypothesis"" posited in the CGEL
- discuss some approaches (the corpus-based approach of the Collins COBUILD dictionaries)
- present some on-line resources and techniques
- few words about parsing and PoS-tagging (the quest for head nouns)

common misconceptions

I will first sum up some of the widespread assumptions I have come across about NCCs. A review of the literature seems to suggest that :

1. potential head nouns of NCCs form a limited class of nouns
2. these nouns have to be 'abstract' nouns
3. head nouns are usually singular
4. determined by *the*
5. the head noun is more often than not the subject of the main clause in the examples given
6. there is always *that*
7. the *that*-clause is mostly adjacent to the noun

Agreg-like question : the derivational hypothesis

- sample (CGEL: 965)
- semantic types
- suggested properties (morphological derivation??)

the *dictum* / *modus* approach

(Ballier 2012, Kanté 2013, Kanté 2016)

(19) Erm what I'd like to er start off with is to emphasize why you're here and erm the fact that I'm not going to patronize you as a driver. (BNC, FM1)

" there is a factive mood whose use in a sentence implies that the speaker is certain that the sentence is true, and a range of other moods indicating hearsay, doubt, and other judgements of the speaker about the sentence. While this distinction is not overt in English, it seems to us that it may be sensed in an ambiguity of declarative sentences. (Kiparsky & Kiparsky 1970: 367)

logophoricity scale (Culy 1994)

The puzzle lay in his certainty that his feeling for her found an echo in her own feelings. (BNC, FP1 : 834)

speech >> thought > > knowledge >> direct perception (Culy 1994 : 1062).

head noun as trigger (*modus*) and a logophoric domain of discourse (*dictum*) : "stretches of discourse in which a person's words, thought , knowledge, or emotions are being reported" (Culy 1994: 1057). CULY, C., 1994 "Aspects of Logophoric Marking", *Linguistics* 32, 1055-1094.

OTHER APPROACHES

Sinclair and Fox, 1990, *Collins Cobuild English Grammar*, Collins, 331

nouns used with reported clauses 7.84 There are many nouns, such as ‘statement’, ‘advice’, and ‘opinion’, which refer to what someone says or thinks. Many of the nouns used in this way are related to reporting verbs. For example, ‘information’, is related to ‘inform’, and ‘decision’ is related to ‘decide’. These nouns can be used in report structures in a similar way to reporting verbs. They are usually followed by a reported clause beginning with *‘that’*. He referred to Copernicus’s statement that the earth moves around the sun. He expressed the opinion that Kitchener should be made War Minister. There was little hope that he would survive. Here is a list of nouns which have related supporting verbs and which can be used with *‘that’*-clauses:

A REMINDER

A norm is a pattern of ordinary usage in everyday language with which a particular meaning or implicature is associated. A pattern consists of a valency structure (see below), together with sets of preferred collocations. (Hanks 2013: 92)

A REMINDER on ontologies

"The verb *fire*, for example, in one of its senses, is associated with a set of lexical items denoting *firearms*: people fire guns, rifles, pistols, revolvers, machine guns, Kalashnikovs, and so on. A group of words like this is said to constitute a lexical set in relation to the verb *fire*. The lexical set is united by a common semantic type—namely, they are all firearms. A lexical set of this kind is given a name—the name of the unifying semantic type—which is conventionally written in double square brackets with initial capital letters, thus: [[Firearm]], [[Physical Object]]. Semantic theorists have tried to build semantic types into hierarchical structures called ontologies, as discussed in Hanks and Ježek 2010. Thus, a [[Firearm]] is a [[Weapon]] is an [[Artifact]] is a [[Physical Object]], and so on. This lexical set and its semantic type activate a contrast with other uses of the verb *fire*: for example, '[[Human]] fire [[Human]]' (meaning 'dismiss from employment') and '[[Human]] fire up [[Machine]]' (meaning 'start')." (Hanks 2013: 12-13)

The generative lexicon: the generative approach (Johnston and Busa 1996; Pustejovsky 1995)

Fig. 1. An example of an enumerative entry for noun *knife*.

$$\begin{bmatrix} \textbf{knife} & \\ \text{TYPESTR} = & \begin{bmatrix} \text{ARG1} = \boxed{\text{x}}\ \textbf{artifact tool} \end{bmatrix} \\ \text{ARGSTR} = & \begin{bmatrix} \text{D-ARG1} = \boxed{\text{y}}\ \textbf{physical object} \\ \text{D-ARG2} = \boxed{\text{w}}\ \textbf{human} \\ \text{D-ARG3} = \boxed{\text{z}}\ \textbf{human} \\ \text{D-E1} = \boxed{e_1}\ \textbf{transition} \\ \text{D-E2} = \boxed{e_2}\ \textbf{process} \end{bmatrix} \\ \text{QUALIA} = & \begin{bmatrix} \text{FORMAL} = \boxed{\text{x}} \\ \text{CONSTITUTIVE} = \{\textbf{blade}, \textbf{handle}, \ldots\} \\ \text{TELIC} = \textbf{cut act}(\boxed{e_2}, \boxed{\text{w}}, \boxed{\text{x}}, \boxed{\text{y}}) \\ \text{AGENTIVE} = \textbf{make act}(\boxed{e_1}, \boxed{\text{z}}, \boxed{\text{x}}) \end{bmatrix} \end{bmatrix}$$

Fig. 2. An example of a generative entry for noun *knife*.

Valency Grammar : Herbst et al. 2004

claim *noun*

P1 **A** Many Hungarians refuse to believe they could have been overtaken and dispute the validity of the statistics used to back up such *claims*.

B According to federal officials, the insurance program has more than enough funds in reserve to pay all *claims*.

P2 **+ to-INF**

A Despite their academic backgrounds, these four women make no *claim* to give us the music as it might have been heard.

P3 **+ that-CL** *(very frequent)*

A The Prime Minister furiously denied Labour *claims* that the Tories are planning massive increases in VAT if they win the forthcoming election. ▪ Judges rejected a defence *claim* that prosecution scientists had deliberately withheld crucial evidence.

P4 **+ about N/V-ing / about N V-ing / about wh-CL**

A The National Union of Teachers has been accused of scaremongering over a *claim* about the influence of Militant Tendency. ▪ Once again Peter Lilley has made exaggerated *claims* about reducing waste in the benefit system. ▪ Classrooms are more overcrowded now than they have ever been despite Mr Clarke's *claim* about more money being spent on education. ▪ There are still conflicting *claims* about how the incident started and who was responsible for setting off the violence.

P5 **+ against N**

B The Mosaic law not only established procedural guarantees before the law, but granted the powerless certain economic *claims* against the wealthy.

P6 **+ by N**

A Ian Brady is to sue a newspaper over allegations made by a woman who once believed she was his daughter. The *claims* by Christine Hart, 30, appeared in the Sunday Express.

B Sinead Mulhern, chief legal officer to the Equal Opportunities Commission, said the settlement could encourage *claims* by women who were sacked after less than two years in a job.

P7 **+ for N/V-ing** *(frequent)*

A The authors make ambitious *claims* for their findings.

B Her *claim* for $500,000 has been rejected and she has, instead, been offered $ 50,000 towards legal expenses. ▪ We had to turn down his *claim* for repainting the windows, but if he had had an accidental damage extension to his policy we would have met the cost.

noun features ?

that-CL	clause introduced by *that* (with adjectives and nouns often replaced by an *if*-clause in conditional contexts) It is akward that a third party is taking a close interest in the reorganisation.
(that)-CL	*that*-clause in which *that* can be omitted *Remember* that the finest cuisines in the world are based on the sauce. *Remember* no alcohol is sold or can be consumed here.

##WordNet (Miller et al. 1990; Fellbaum 1998])

- a computational lexicon of English
- based on psycholinguistic principles
- about 155,000 words organized in over 117,000 synsets. (Wordnet 3.0)

'We can view a synset as a set of word senses all expressing (approximately) the same meaning' (Navigli 2009)

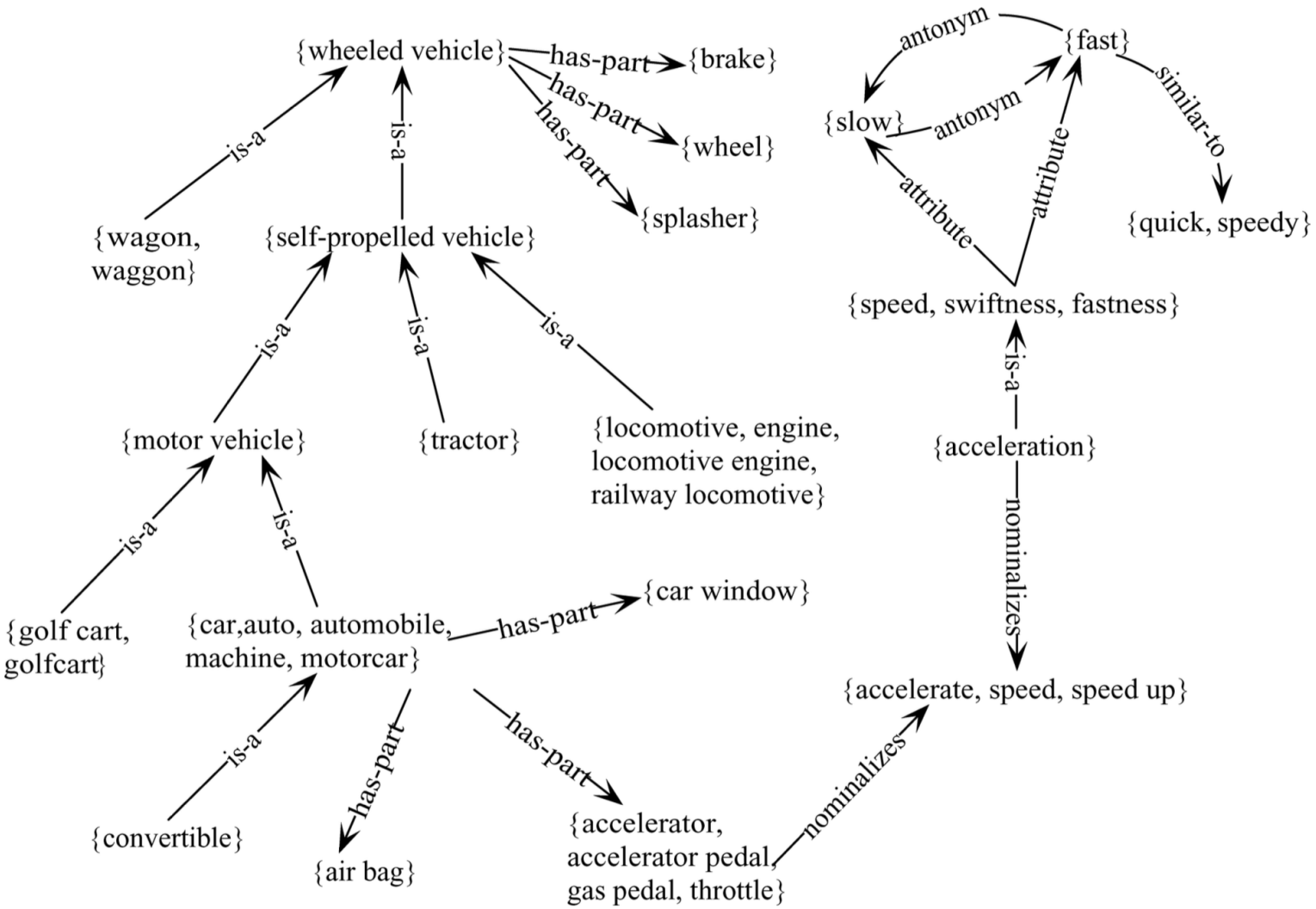
A SYNSET, an example from (Navigli 2009)

$$
\begin{gathered}
Senses_{WN}(car_n) = \{\{car_n^1, auto_n^1, automobile_n^1, machine_n^4, motorcar_n^1\}, \\
\{car_n^2, rail\ car_n^1, rail\ way\ car_n^1, rail\ road\ car_n^1\}, \\
\{cable\ car_n^1, car_n^3\}, \\
\{car_n^4, gondola_n^3\}, \\
\{car_n^5, elevator\ car_n^1\}\}.
\end{gathered}
$$

Synsets (Wordnet)

{*act, action, activity*}
{*animal, fauna*}
{*artifact*}
{*attribute, property*}
{*body, corpus*}
{*cognition, knowledge*}
{*communication*}
{*event, happening*}
{*feeling, emotion*}
{*food*}
{*group, collection*}
{*location, place*}
{*motive*}
{*natural object*}
{*natural phenomenon*}
{*person, human being*}
{*plant, flora*}
{*possession*}
{*process*}
{*quantity, amount*}
{*relation*}
{*shape*}
{*state, condition*}
{*substance*}
{*time*}

The Semantic network



Semantic relations (Wordnet)

Table 1. Semantic Relations in WordNet

Semantic Relation	Syntactic Category	Examples
Synonymy (similar)	N, V, Aj, Av	pipe, tube rise, ascend sad, unhappy rapidly, speedily
Antonymy (opposite)	Aj, Av, (N, V)	wet, dry powerful, powerless friendly, unfriendly rapidly, slowly
Hyponymy (subordinate)	N	sugar maple, maple maple, tree tree, plant
Meronymy (part)	N	brim, hat gin, martini ship, fleet
Troponomy (manner)	V	march, walk whisper, speak
Entailment	V	drive, ride divorce, marry

assumption in Wordnet: http://wordnetweb.princeton.edu/perl/webwn

- S: (n) **assumption** (the act of assuming or taking for granted) *"your assumption that I would agree was unwarranted"*
  - *direct hyponym* / *full hyponym*
    - S: (n) position (the act of positing; an assumption taken as a postulate or axiom)
  - *direct hypernym* / *inherited hypernym* / *sister term*
    - S: (n) act, deed, human action, human activity (something that people do or cause to happen)
      - S: (n) event (something that happens at a given place and time)
        - S: (n) psychological feature (a feature of the mental life of a living organism)
          - S: (n) abstraction, abstract entity (a general concept formed by extracting common features from specific examples)
            - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
  - ***derivationally related form***
    - W: (v) assume [Related to: assumption] (take to be the case or to be true; accept without verification or proof) *"I assume his train was late"*

beyond the remit of agreg questions?

- MWUs ?
- some parsing issues

Multi-Word Units / MWE ?

The PHRASEME project: the grammatical tagging of nominal Multi-Word Units: parsing texts and nominal expressions for PoS: (*an over-ther-counter*) (more about this and comparable constructions, Ballier 2016)

'in the sense that'/ 'on the grounds that' / 'on the basis that

tenor of the cross-examination / * word of honur*?

ESP uses?? *the rider that* (l'avenant, la clause additionnelle, le codicille), *oversight, off chance, applications, direction, practice, guilt, submission*

SHALLOW *versus* DEEP PARSING

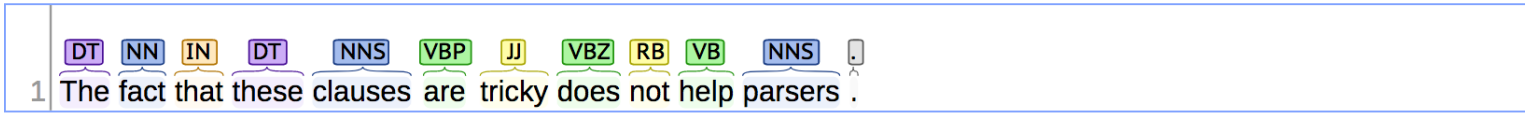
- parsed corpora : ICE-GB, WSJ
- parsers
- parsing issues
- the ontology of tagsets (Penn Treebank,)
- CLAWS5 vs. CLAWS7: *that*
- head nouns in Acquis Communautaire (Zelenakova 2015 on precision and recall)
- querying a corpus with PoS tags (Corpus Workbench, Antconc for PoS)

PoS-tagging and parsing

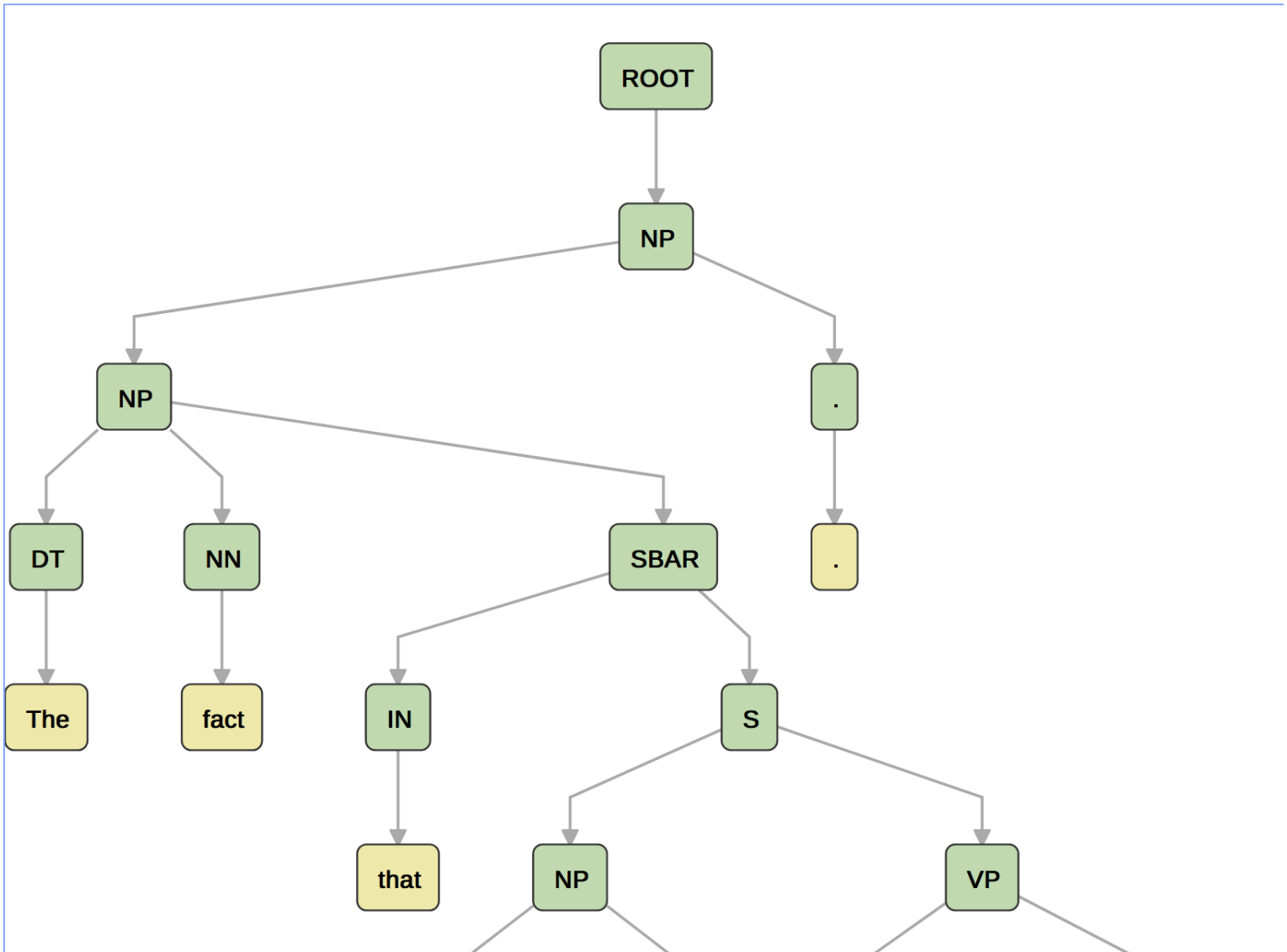
- using parsers for parsed data (ICE-GB...)
- TregEx

##ON-LINE PARSERS : one example of PoS-tagging and Parsing

Stanford's CoreNLP : http://corenlp.run/



Constituency Parse:



NEXT PLANS

Replicating (Price et al., 2008): extending the inventory of head nouns

Word Embeddings (Mikolov et al. 2013)

the Word2vec algorithm and cosine distance to *"understanding"*

```
Word: understanding  Position in vocabulary: 1475
```

Word	Cosine distance
knowledge	0.610022
understand	0.549965
nature	0.548277
intuition	0.516479
insights	0.508795
insight	0.508529
subjectivity	0.508239
thinking	0.505278
ideas	0.490778
rationality	0.487866
implications	0.486571
idea	0.483600
enquiry	0.474852
conception	0.473498
experience	0.473029
interpretation	0.472069
understandings	0.463555
explanation	0.463139
concepts	0.461290

Behavioural profiles (Levshina 2015: 304)

- Behavioural profiles, cf. Vergaro (to appear) in Arigne & Migette (2018)
- clustering nouns with similar constructions (cf. verbs with causative constructions, Gries and Divjak)
- disentangling IDEA (*that*, *to*) and other heads nouns

TAKE HOME MESSSAGE(S)

- from periphery to the core "A Lexically Based Approach to Linguistic Theory" (Hanks 2013)
- verb-inspired conceptualisations of nouns (argument structure: Grimshaw 1990, "complementation"?)
- "The kernel of generative semantics was an obliteration of the syntax-semantics boundary at the deepest level of grammar – the axiom that the true deep structure was the semantic representation, not a syntactic input to the semantic component". (Harris 1993 : 105)
- "constructicon" and enunciativist research agenda "You shall know a word by the company it keeps" (Firth) and Lapaire & Rotgé 1991 "affinités électives"
- more technical explorations of this "company" in the pipeline
- useful resources to explore for the agreg : Hank 2013, Herbst et al. 2004, Wordnet (Miller 1995)

SOME REFERENCES

Arigne, V. & Chr. Rocq-Migette (eds.) 2018 *Theorization and Representations in Linguistics*. Newcastle: Cambridge Scholar Publishing.\ Ballier, Nicolas. (2007). La complétive du nom dans le discours des linguistes. In *La Coordination et la subordination dans le texte de spécialité, sous la direction de David Banks. Paris: l'Harmattan, pp. 55-76.\ Ballier, Nicolas (2010) « La complétive du nom, un "dictum" pris en charge par personne ? », dans* Les représentations linguistiques de la personne, *José Antonio VICENTE LOZANO, Alain BLANC et Nicolas BALLIER (éds.), Collection linguistique Épilogos, 1, Rouen, Publications Électroniques de l'ERIAC, p. 213-234.\ http://eriac.univ-rouen.fr/la-completive-du-nom-un-dictum-pris-en-charge-par-personne \ Ballier, N. (2015). "Phrasal compounds" and the discourse/lexicology interface:"conglomeration" within the French tradition of English lexicology. SKASE Journal of Theoretical Linguistics, 12(3), 115–142.\ Culy, Christopher (1994): "A Note on Logophoricity in Dogon", in* Journal of African Languages and Linguistics* 15:2: 113-125.\

REFERENCES (cont.)

Hanks, Patrick, and James Pustejovsky. 2005. "A pattern dictionary for natural language processing." *Revue française de linguistique appliquée* 10 (2).\ Herbst, Thomas et al. (2004). *A Valency Dictionary of English: A Corpus-Based Analysis of the Complementation Patterns of English Verbs, Nouns and Adjectives*. Walter de Gruyter.\ Kanté, Issa (2017) *Étude sémantico-syntaxique de la complétive nominale en anglais et en français*. Collection Sémantiques. Paris : L'Harmattan. \ Kanté, Issa (2010) « Mood and modality in finite noun complement clauses: A French-English contrastive study ». In M. Stefania; K. Heylen and G. De Sutter (eds.), *International Journal of Corpus Linguistics*. Corpus Studies in Contrastive Linguistics. 15: 2. Amsterdam: John Benjamins, pp. 267-290.\ Keizer, E. (2007). *The English noun phrase: The nature of linguistic categorization*. Cambridge University Press.\ Miller, G. (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41. https://pdfs.semanticscholar.org/ec6e/56339e0b2f5c2f464a7a48b536b78bfeafba.pdf

THANK YOU

nballier AT free DOT fr