

Aim

Ongoing project of lexical resources of French cross-disciplinary scientific lexicon in the framework of the Termith ANR project (<http://www.atilf.fr/ressources/termith/>)

- Freely available
- Specific lexicon: deals with words which cannot be used as terms but refer to methods, arguments, opinions and metatext, for example *hypothesis, that is why, surprisingly ...* (for English, see Kosem 2010; Granger & Paquot 2010)

COLLOCATIONS

- ✓ In line with Explanatory and Combinatorial Lexicology (Mel'čuk, 1998), collocations are compositional and prefabricated binary multiword expressions.
- ✓ Each component will receive a semantic tag and be stored within the database entry (e.g. s.v. *hypothèse* for *réfuter une hypothèse* ('to disprove a hypothesis') or *hypothèse valide* ('valid hypothesis').
- ✓ Fully frozen MWE expressions (e.g. *mettre en évidence* 'highlight') will be stored as single units.

Definition

SYNTACTIC TYPES OF COLLOCATIONS

- ✓ **N prep N** : *hypothèse de travail* (work hypothesis), *formulation d'une hypothèse* (formulation of hypothesis)
- ✓ **N Adj** : *hypothèse valide* (valid hypothesis)
- ✓ **N V** : *les résultats confirment ...* (results confirm...)
- ✓ **V (prep) N** : *faire une hypothèse* (to make a hypothesis...)
- ✓ **V ADV** : *assumer pleinement* (fully assume)
- ✓ **Adv Adj** : *totalemment absent* (totally absent, *significativement différent* (significantly different)

COLLOCATION EXTRACTION

- ✓ **Corpus** : 500 research articles in 10 disciplines of human sciences
- ✓ Core **cross-disciplinary scientific lexicon** of ~1200 words (Hatier 2013)
- ✓ **Syntax-based extraction of collocations** (Seretan 2011) with the help of the **Lexicoscope** (Kraif & Diwersy 2014)
- ✓ Several statistical **thresholds** are used: frequency over 7 occurrences, log-likelihood ratio > 10.7 and a dispersion in at least 3 disciplines.
- ✓ Manual checking of results

Method

Le lexicoscope (Kraif & Diwersy 2014) (<http://emolex.u-grenoble3.fr/emoBase>)

Example : extraction of V-obj-*hypothèse* collocations

Pivots (lemme Forme): Catégorie: Traits:

Relations:

Définition du contexte des pivots:

Concordances Cooccurrences

I1	I2	f.deprels	f	f1	f2	N	f.disp	am.log.likelihood	r.log.likelihood
hypothèse_*	faire_VERB	-OBJ	120	571	8939	280230	10	270,1137	1
hypothèse_*	tester_VERB	-OBJ	38	571	239	280230	8	246,6283	2
hypothèse_*	émettre_VERB	-OBJ	20	571	116	280230	8	142,2757	3
hypothèse_*	rejeter_VERB	-OBJ	23	571	213	280230	4	140,8877	4
hypothèse_*	avancer_VERB	-OBJ	20	571	186	280230	9	128,9839	5
hypothèse_*	formuler_VERB	-OBJ	17	571	135	280230	7	109,4272	6
hypothèse_*	valider_VERB	-OBJ	13	571	115	280230	7	80,8485	7
hypothèse_*	confirmer_VERB	-OBJ	19	571	478	280230	5	78,1058	8
hypothèse_*	vérifier_VERB	-OBJ	14	571	280	280230	5	65,7744	9
hypothèse_*	accepter_VERB	-OBJ	10	571	296	280230	3	37,8284	10
hypothèse_*	poser_VERB	-OBJ	9	571	705	280230	5	18,0858	11

PERSPECTIVES: INFORMATION ON SYNTACTIC PROPERTIES OF COLLOCATIONS

Collocation	Faire-hypothèse	Exemples
Determiners	I' (91%), des (8%), une (0,5%)	
Syntactic alternations	Direct objet	92%
	Passive	2%
	Reduced passive	5%
External subcategorization	Que-P	(with I') : 58%
	Prep-de	(with I') : 12%

Examples for 'Faire-hypothèse':
Nous avons fait l'hypothèse
Ainsi, aucune hypothèse n'est faite sur le caractère ...
les conséquences des hypothèses générales faites sur l'illocutoire
On peut donc faire également l'hypothèse que c'est au cours de la Belle Epoque que s'opère la transition
On peut faire l'hypothèse d'une forte homogénéité des sujets ...

References

- Evert, S. (2008). Corpora and collocations, in Anke Lüdeling/Merja Kytö (eds.), *Corpus Linguistics. An International Handbook*. Berlin: Mouton/de Gruyter, 1212-1248.
- Hatier, S. (2013). Extraction des mots simples du lexique scientifique transdisciplinaire dans les écrits de sciences humaines : une première expérimentation. Dans *Actes de Recital/2013* (p. 138-149). Les Sables d'Olonne, France.
- Kraif, O., & Diwersy, S. (2014). Exploring combinatorial profiles using lexicograms on a parsed corpus: a case study in the lexical field of emotions. In P. Blumenthal, I. Novakova, & D. Siepmann (eds.), *Actes Du Colloque International Nouvelles Perspectives En Sémantique Lexicale Et En Organisation Du Discours*. Osnabrück, Allemagne: Peter Lang.
- Mel'čuk, I. (1998). Collocations and Lexical Functions. In A. P. Cowie (ed), *Phraseology. Theory, Analysis and Applications*, 23-53. Oxford : Clarendon Press.
- Seretan, V. (2011). *Syntax-based collocation extraction*. Vol. 44. Springer Science & Business Media, 2011.
- Tran, T. T. H. (2014). Les séquences lexicalisées à fonction discursive comme outil d'aide à l'écriture auprès des étudiants étrangers. Dans P. Blumenthal, I. Novakova, & D. Siepmann (éd.), *Actes Du Colloque International Nouvelles Perspectives En Sémantique Lexicale Et En Organisation Du Discours*. Osnabrück, Allemagne: Peter Lang.
- Tutin, A. (2014). La phraséologie transdisciplinaire des écrits scientifiques : des collocations aux routines sémantico-rhétoriques. Dans A. Tutin & F. Grossmann (éd.), *L'écrit Scientifique : Du Lexique Au Discours. Autour De Scientext* (p. 27-44). Rennes: Presses Universitaires de Rennes.