Towards an automatic measurement of verbal lexicon acquisition: the case for a young children-vs-adults categorization in French and Mandarin ⋆

Yann Desalle[a], Shu-Kai Hsieh[b], Bruno Gaume[c], and Hintat Cheung[b]

[a] Octogone, Université de Toulouse,Toulouse, France
yann.desalle@univ-tlse2.fr
[b] Graduate Institute of Linguistics, National Taiwan University, Taipei, Taiwan
{shukai, hintat}@ntu.edu.tw
[c] CLLE-ERSS, CNRS - Université de Toulouse, Toulouse, France
gaume@univ-tlse2.fr

Abstract. In this paper we define a lexical metrology in graphs of verbal synonymy to compute the *flexsemic* score of speakers from their verbal productions in action denomination tasks. This flexsemic score is used to automatically categorize young children versus young adults. We show that this score is effective in French and in Mandarin.

Keywords: automatic categorization, Flexsem, language acquisition, Prox

1 Introduction

This paper focuses on the automatic measurement of verbal lexicon acquisition. Our research combine two sides: a psychological side and a computational side. In a recent work (Gaume *et al.*, 2008) we showed that the psychological results on verbal lexicon acquisition fit our computational model of the semantic organization of the verbal lexicon. Based on these results, we propose here to use our computational model to build a measurement of the lexical acquisition useful to categorize the healthy young children vs. healthy adults.

The first part of this paper presents the Approx protocol, the data collected in French and Mandarin and the first psychological results about verbal lexicon acquisition. Then, the second part accounts for the computational model and describes three preliminary measurements useful in our categorization while the next section describes how to assign a "flexsemic" score to a participant. Then, the way we used this score for the categorization task and the results analysis are exposed. This paper ends with a short presentation of the web platform Flexsem usable to build online categorizations.

2 Approx protocol

2.1 General overview

In order to compare the lexical acquisition of verbs in French and Mandarin, we used the Approx protocol initially built to study the production of verbal approximations and the hierarchical dimension of the lexicon in normal and pathological development (Duvignau *et al.*, 2005; Gaume *et al.*, 2008). The material of this protocol consists in 17 speechless action-movies divided as in Table 1 into 3 action concepts: TO DETERIORATE, TO REMOVE, TO SEPARATE

In each movie, a woman alters an object from an initial state to a final state with the help of her hands or with an instrument.

⋆ The research presented in this paper is supported by the ANR-NSC (France-Taiwan) bilateral project M3 (Modeling and Measurement of Meaning).

Table 1: The 17 action movies of the Approx protocol

TO DETERIORATE	TO REMOVE	TO SEPARATE
Burst a balloon	Peel a carrot	Make bread-crumbs
Crumple a piece of paper	Peel an orange	Cut a bread
Break a glass	Pull the bark of a log off	Break a bread off
Crush a tomato	Undress a doll	Chop parsley
Tear off a newspaper	Take down legos	Saw a plank of wood
	Peel a banana	Unsew a shirt

These movies were shown to several participants. For the M3 project, we focused on two populations in French and Mandarin : healthy children and healthy adults.

2.2 Participants

French participants At the time being, 170 children between 2 and 11 years old and 75 young adults between 18 and 40 years old have passed the Approx protocol . Children were recruited from kindergartens and elementary schools located in Toulouse, France.

Mandarin participants In Mandarin, 60 children between 2 and 9 years old and 60 young adults old (students) have passed the Approx protocol . Children were recruited from kindergartens and elementary schools located in Taipei, Taiwan.

2.3 Procedure and verbs extractions

Procedure Before showing the video, participants were told what the ”lady” (or the ”big sister” in Mandarin when participant are children) was going to perform. After showing the video, the experimenter asked the children what the woman had done. Between each action movie, a distractor (i.e. popular cartoon characters in Mandarin Team) was shown to the children to avoid perseveration effects. Children who didn’t finish watching the 17 movies were discarded.

Verbs extractions Verbal forms that denominate the action were extracted from each elicited response and were lemmatized. When a verbal form was compound, it was split, depending on its components, into:

- simple verb + complement (e.g. *to break into pieces* → *to break* + *into pieces*)
- simple verb + simple verb (e.g. *to make broken* → *to make* + *to break*)
- simple verb + result when the verb is a mandarin resultative compound verb (Thompson, 1973; Lu, 1977; Gu, 1992; Cheng and Huang, 1994; Gao, 1997) (it is not useful in French)

Like this, the verbal forms recorded in French and Mandarin were built in the same way, and so, one can compare the analyses of the verbal productions between the two languages.

2.4 First psycholinguistic results

A first study, based on the Approx data, has been done in French to compare the verbal productions of the healthy 2-4 years old children and the adults who have stabilized their lexicon structure. The verbal productions were compared on two criterion:

- the verbs’ genericity to discriminate, in front of the denominated action, the specific verbs (e.g. *to saw* / SAWING A PLANK OF WOOD) from the generic verbs (e.g. *to cut* / SAWING A PLANK OF WOOD);

- the verbs' conventionnality to discriminate the conventional verbs that belong to the semantic field of the action and fit the action (e.g. *to peel* / PEELING A BANANA), the intra-domain approximations that belong to the semantic field but not well fit the action (e.g. *to cut* / PEELING A BANANA) and the inter-domain approximations that are out of the semantic field of the action but convey the action's concept (e.g. *to undress* / PEELING A BANANA convey the concept /TO REMOVE/).

All the data analyses come from the agreement of two french linguists.

The results confirm that 2-4 years old children produce more approximations than adults (Table 1). Each children produced between 2 and 5 approximations.

Table 2: Percentages of approximations produced with Approx in French

	CHILDREN	ADULTS
Percentage of approximations	34%	5%
Percentage of intra-domain approximations	24%	4%
Percentage of inter-domain approximations	10%	1%

According to our hypothesis, these differences between young children and adults might be retrieved by the study of the lexical organization of verbs.

Thereafter, in this article, we only focus on young children (between 24 and 59 months) and adults (between 18 and 40 years) in French and Mandarin.

3 Measurement of semantic proximity in French and Mandarin

3.1 A computational approach

Since Zipf (Zipf, 1949), many researchers tended to build theories of language that account for language statistical regularities. Recently, Ferre-i-Cancho (Ferrer-i-Cancho and Sole, 2001) and Sigman (Sigman and Cecchi, 2002) who work on graph theory highlighted the "hierarchical small world" (HSW) structure used in language studies. From our part, we use the general framework of HSW in a specific way:

- a new linguistic and psycholinguistic insight that guides us and help us on our results validation;
- the kind of objects studied (dictionaries);
- our analysis of graph structure resulting in a computational model of semantic proximity among vertices (here vertices are verbs).

The computational model for verbs similarity we built in this way handles metaphorical pairs such as *to undress* / *to peel* (Desalle *et al.*, 2009; Gaume *et al.*, 2002; Gaume *et al.*, 2010; Sajous *et al.*, 2010) contrary to other existing models (Resnik and Diab, 2000). It is why we aim to use it to catch the approximations produced by Approx participants.

3.2 Hierarchical Small Word Networks

Thereafter, a graph $G = (V, E)$ is defined by a set V with n vertices and a set $E \subset V^2$ with m edges.

In this paper

- V is a set of words;
- E is defined by the relation $V \overset{R}{\longmapsto} V : (u, v) \in E \iff u \overset{R}{\longrightarrow} v$

- $deg(u)$ is the degree of the vertex u in G (the number of neighbors of u):

$$deg(u) = |\{v \in V/(u, v) \in E\}|$$

Most of the lexical networks, as networks extracted from real world, are Hierarchical Small Worlds (HSW) networks:

- Edge sparsity: HSW are sparse in edges $m = O(n \log(n))$;
- Short paths: in HSW, the average path length (**L**)[1] is short. Most of the time, there is at least one short path between two vertices;
- Hight clustering: in HSW, the clustering coefficient (**C**), that is the probability of two distinct vertices that adjacent to a third vertex to be adjacent, is an order of magnitude higher than for Erdos-Renyi (random) graphs: $C_{HSW} \gg C_{random}$; that means that the graph is locally dense, although it is globally sparse;
- Heavy-tailed degree distribution: in HSW graph, the distribution of vertices' degrees fits a Power Law. The probability $P(k)$ of a vertex to have k neighbors decreases as a Power Law, $P(k) \approx k^a$ (a is a constant characteristic of the graph). Random graphs fit a Poisson Law.

3.3 The mathematical model: the fleximity

We consider a walker randomly walking on the vertices of the graph. The probability of the walker to be at a particular vertex is given by the Markov chain on the graph. This Markov chain is defined by the transition Matrix $[G]$ on $V \times V$:

$$[G] = (g_{u,v})_{u,v \in V}, \text{such as } g_{u,v} = \begin{cases} \frac{1}{deg(u)} \textbf{ if } (u, v) \in E, \\ 0 \textbf{ otherwise}. \end{cases} \quad (1)$$

If Δ_0 is the vector in V of the initial distribution probability of the walker (that is the probability to be at any vertex at the time $t_0 = 0$), the vector in V of the probability distribution of this walker after t steps becomes:

$$[C(\Delta_0, t)] = \Delta_0 [G]^t \quad (2)$$

Thus, to measure, in $G = (V, E)$, the proximity between a vertex $s \in V$ and the initial state Δ_0 of the walker, we define $prox(G, \Delta_0, s, t)$ by the t steps random walk as:

$$prox(G, \Delta_0, s, t) = \frac{[C(\Delta_0, t)]_s}{\max_{v \in V}([C(\Delta_0, t)]_v)} \quad (3)$$

For a graph $G = (V, E)$ and an initial state Δ_0, we can now define the flexemity at the time t of a vertex $s \in V$:

$$flex(G, \Delta_0, s, t) = \frac{prox(G, \Delta_0, s, t)}{deg(s)} \quad (4)$$

The s $flex$(emity) is the s $prox$(imity) weighted by the inverse of s degree.

Now, we will present the dictionary graphs on which we apply this model.

3.4 Dictionary graphs

Meaning in dictionary definition is at least partially brought by the relations they create between the words constituting the entries. Our approach consists in exploiting the small word properties of the graphs corresponding to dictionaries. More precisely, we are taking advantage of our hypothesis that clusters correspond to areas of closely related meanings. To illustrate our approach we use a French dictionary graph of verbs and a Mandarin dictionary graph of verbs.

[1] Average length of the shortest paths between two vertices

French dictionary graph of verbs DicoSyn.Verb is a graph of the verbs extracted from DicoSyn[2]: there is an edge A, B if the verbs described by the vertices A and B are synonyms in DicoSyn. Then DicoSyn.Verb has been given symmetric and reflexive.

Mandarin dictionary graph of verbs CilinCWN.verb is a graph of verbs extracted from CilinCWN.verb: a fusion of Chinese Wordnet (CWN)[3] and a Chinese thesaurus TongYiCi CiLin (Cilin)[4] In order to compare with French graph, data have been preprocessed following DicoSyn.Verb. More details can refer to (Gaillard *et al.*, 2010).

Table 3: Properties of DicoSyn.Verb and CilinCWN.verb on the largest connected component

	DicoSyn.Verb	CilinCWN.verb
n	8993	8393
m	111659	94316
L	4.19	5.65
C	0.14	0.61
λ	-2.02 (0.93)	-1.79 (0.61)

The table 3 shows the properties[5] of DicoSyn.Verb and CilinCWN.verb on their largest connected component: they are typical Hierarchical small worlds.

4 How to assign a "flexsemic" score to Approx participants

4.1 General overview

We aim to automatically categorize the *24-59* months old children versus *18-40* years old adults in French and Mandarin. To do it for a language L, we used the verbal graph G_L[6] and three measurements : deg, $prox$ (equation 3), and $flex$ (equation 4).

The categorization is based on the verbs produced by the participants. But, whereas the verbs' degrees depend only on the graph G_L, the $flex$(emity) and the $prox$(imity) measurements need, beforehand, the initial distributions of the verbs.

4.2 The initial distribution

The verbs depend on the action movie they denominate. So, one initial distribution on G_L per movie is to be taken into account per movie. We firstly identified two ways to define these distributions:

- A linguist gives a verb v per movie, named the referent verb for this movie. Therefore, for this movie the initial distribution $\Delta_0 = \delta_v$ where δ_v is the certitude to be on v;

[2] Dicosyn is a compilation of synonym relations extracted from seven other dictionaries (Bailly, Benac, Du Chazaud, Guizot, Lafaye, Larousse et Robert). Dicosyn has been first realized at ATILF (Analyse et Traitement Informatique de la Langue Francaise), before being corrected at CRISCO laboratory (http://elsap1.unicaen.fr/dicosyn.html) (Ploux and Victorri, 1998)

[3] Chinese Wordnet is a lexical resource modelled on Princeton WordNet, with many novel linguistic considerations for Chinese. It is proposed and launched by Huang et al. (Huang *et al.*, 2004), at the time of writing it contains 28,815 synonyms

[4] The Tongyici Cilin (Mei *et al.*, 1984) is a Chinese synonym dictionary known as a thesaurus in the tradition of Roget's Thesaurus in English. It contains about 70000 lexical items under 12 broad semantic classes marked from A to L. These broad classes are further divided into 94 subclasses, and 1,428 heads. But in our experiment, classes of A-E and L are removed, for they refer to non-verbal entities like human, physical object, time and space, features, etc.

[5] n: number of vertices, m: number of edges, L: average path length, C: clustering coefficient, λ: slope of the power-law model (in parenthesis: the correlation coefficient of this model)

[6] $DicoSyn.Verb$ for French and $CilinCWN.verb$ for Mandarin

- For each action movie, the most frequent verb produced by the adults in Approx is the referent verb. The initial distribution is defined as just above.

But considering our goal and the populations to categorize, these propositions are methodologically biased. Indeed, the adults (and the linguists are adults) have to be categorized, so we cannot take a referent verb from the adults' verbs only [7]: if we do it, $prox$ and $flex$ measurement might highlight a difference between the two populations except if the adults and the children produced the same verbs for each movie.

Nevertheless, there is another way to build a referent, without methodological bias, and based on a list of verbs rather than only one verb: a list of verbs fairly and randomly extracted from the adults' (25) and the children's (25) verbs is given has a referent set for an action movie m[8]. Then, given a referent set of verb U and $freq(v)$ the frequency of the verb v in U, the initial distribution Δ_0 is defined such as:

$$\begin{cases} [\Delta_0]_v = \frac{freq(v)}{\sum_{s \in U} freq(s)} \textbf{if } v \in U \\ [\Delta_0]_v = 0 \textbf{ otherwise}. \end{cases} \tag{5}$$

As the initial distribution Δ_0 is now defined for each movie, the next section will explain how the three measurements are used to categorize populations.

4.3 Categorization of young children vs. adults

Aftewards, the number of steps t of all the random walks is 4.

Given an Approx participant p, and a graph $G_L = (V, E)$, we define 3 scores, $Degree_{score}$, $Prox_{score}$ and $Flex_{score}$, as below[9] :

- $Degree_{score}(p)$ is the average of the degrees of the verbs produced by the participant p in the graph;
- $Prox_{score}(p)$ is the average, on the set of movies, of the $prox$(imities) of the verbs produced by the participant p from the referent set.
- $Flex_{score}(p)$ is the average, on the set of movies, of the $flex$(emities) of the verbs produced by the participant p from the referent set.

In this way, it is possible to assign a "degree" score, a "proxemic" score and a "flexsemic" score to each participant.

Our method is based on the observation of the psycholinguistic Approx results (see §2.4): children produce more approximations and more generic verbs than adults.

Therefore, our hypotheses are that:

- the difference in the number of generic verbs between a child p_c and an adult p_a fits a difference in the $Degree_{score}$ values, as below:

$$Degree_{score}(p_c) > Degree_{score}(p_a) \tag{6}$$

- the difference in the number of approximations between a child p_c and an adult p_a fits a difference in the $Prox_{score}$ values, as below:

$$Prox_{score}(p_c) < Prox_{score}(p_a) \tag{7}$$

---

[7] The bias is higher with the verbs frequently produced by adutls

[8] one verb can occur many times in the set if it is produced by many participants.

[9] If a verb v is compound verb that is not in G:

- the $degree$ of v is equals to the lowest $degree$ of its verbal components
- the $prox$(imity) of v from a referent set is equals to the highest $prox$(emity) of its verbal components
- the $flex$(emity)of v from a referent set is equals to the highest $flex$(emity) of its verbal components

And then, if p_c is a child participant and p_a and adult participant

$$Flex_{score}(p_c) < Flex_{score}(p_a) \quad (8)$$

To evaluate the relevance of these scores, we used them for an automatic categorization task run in French and Mandarin, as presented in the next section.

5 Automatic categorization of young children versus adults

To remind, many verbs produced by children are verbal approximations (intra-domain or inter-domain), generic verbs or both of them. So:

- $Degree_{score}$ might success to catch generic verbs but sould fail to catch approximations;
- $Prox_{score}$ might success to catch approximations but should fail to catch generic verbs;
- $Flex_{score}$ might success to catch generic verbs and approximations.

This is what we aim to evaluate here, by describing a categorization task of *24-59 months old* native children versus *18-40 years old* native adults in French and Mandarin (see table 4).

Table 4: Number of native participants in French and Mandarin

	French	Mandarin
24-59 months old	74	28
18-40 years old	76	60

The description below is the same for all the languages.

Description Given a score $S \in \{Degree_{score}, Prox_{score}, Flex_{score}\}$, S was assigned to every young children and adults participants Then, a 2-means[10] clustering on these participants by using S is compared to the original populations (young children and adutls). The results are in the table 5.

Table 5: Results of the 2-means clustering in French and Mandarin

SCORE	LANGUAGE	F-SCORE
$Degree_{score}$	Mandarin	0.65
$Prox_{score}$	Mandarin	0.85
$Flex_{score}$	Mandarin	**0.93**
$Degree_{score}$	French	0.87
$Prox_{score}$	French	0.59
$Flex_{score}$	French	**0.96**

Results These results show that the categorization with "flexsemic" score is the very best. Indeed:

- the categorization with $Degree_{score}$ gives good results in French but not in Mandarin;
- the categorization with $Prox_{score}$ gives good results in Mandarin but not in French;
- the categorization with $Flex_{score}$ gives good results in French *and* in Mandarin with a higher score than $Degree_{score}$ and $Prox_{score}$.

The figures 1 and 2 depict the flexsemic scores per age. The black circles for children and the black squares for adults stand for mis-categorized participants.

[10] We used an euclidian distance.

Figure 1: French flexsemic score acquisition

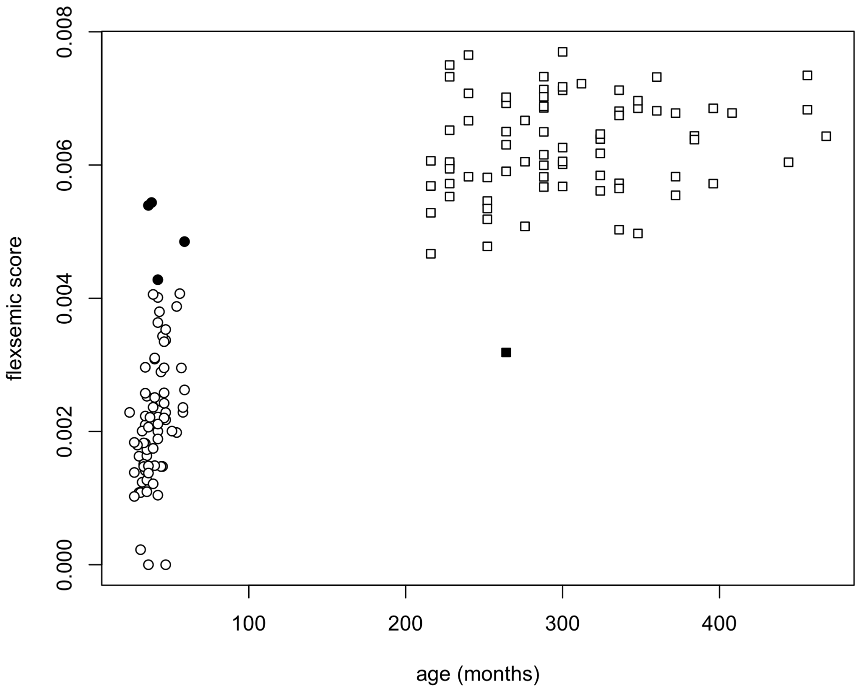
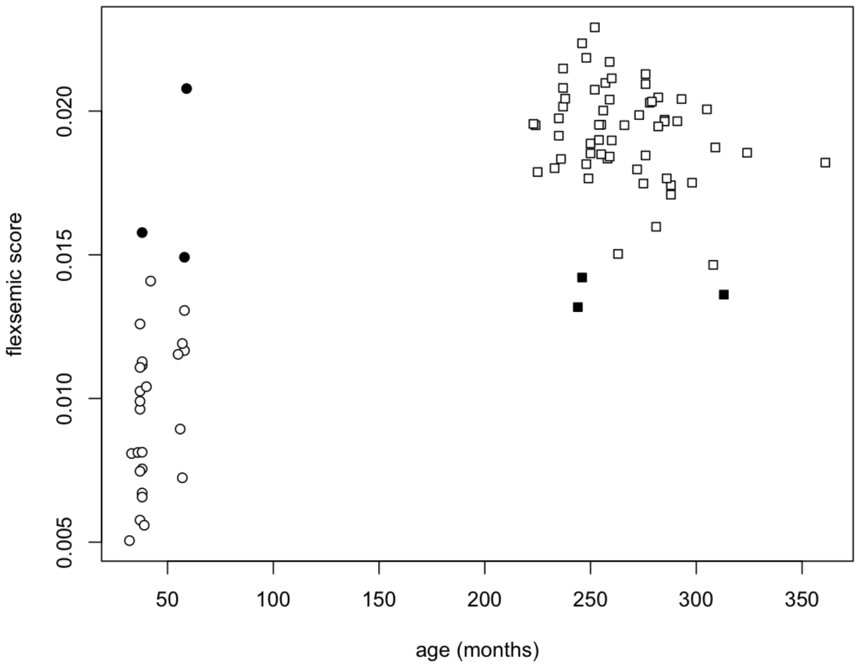


Figure 2: Mandarin flexsemic score acquisition.



6 Flexsem

All the data and all the results presented here are accessible on the web platform Flexsem (http://erss.irit.fr/flexsem). More, in this web platform, the Approx data from many languages (French, Mandarin, Polish, Korean, Portuguese ...), many ages and from pathological populations (Alzheimer, Apserger Autism, Aphasic, Schizophrenics ...) will be recorded . So, the categorization built by using the $Flex_{score}$ on young children and adults should be reused on many populations (selected by combining all the Flexsem's criteria [11]).

7 Conclusion

Based on psycholinguistics observations and on the Hierarchical Small World structure of the lexical graphs we tried to build a new measurement of the acquisition of the verbal lexicon: the "*flexsemic*" score. A first evaluation in French and Mandarin on a categorization task between *24-59 months old* native children and *18-40 years old* native adults and a comparison with other scores showed its relevance. In our future works, on the one hand, the *flexsemic* score will be evaluated in categorization task based on other populations (e.g. *Apserger autism-vs-healty* young children) and, on the other hand, it will be used in predictive profiling applications.

References

Cheng, Lisa Lai-Shen and C.-T. James Huang. 1994. On the argument structure of resultative compounds. In Matthew Y. Chen and Ovid J.L. Tzeng, eds., *In Honor of William S-Y. Wang: Interdisciplinary Studies on Language and Language Change*, pp. 187–221. Pyramid Press.

Desalle, Yann, Bruno Gaume, and Karine Duvignau. 2009. SLAM : Solutions lexicales automatique pour métaphores. *Traitement Automatique des Langues*, 50(1), 145–175.

Duvignau, Karine, Bruno Gaume, and S. Kern. 2005. Semantic approximations intraconcept vs. interconcepts in early verbal lexicon: flexibility against error. In *Proceedings of ELA 2005, Emergence of language abilities: ontogeny and phylogeny*.

Ferrer-i-Cancho, Ramon and Ricard V. Sole. 2001. The small world of human language. In *Proceedings of The Royal Society of London. Series B, Biological Sciences*, volume 268, pp. 2261–2265, November.

Gaillard, Benoît, Yannick Chudy, Pierre Magistry, Shu-Kai Hsieh, and Emmanuel Navarro. 2010. Graph representation of synonymy and translation resources for cross-linguistic modelisation of meaning. In *In Proceedings of Workshop on Models and Measurement of Meaning, at PACLIC conference*, Sendai, Japan.

Gao, Qian. 1997. Resultative verb compounds and ba-construction in chinese. *Journal of Chinese Linguistics*, 25, 84–130.

Gaume, Bruno, Karine Duvignau, O. Gasquet, and M.-D. Gineste. 2002. Forms of meaning, meaning of forms. *Journal of Experimental and Theoretical Artificial Intelligence*, 14, 61–74.

Gaume, Bruno, Karine Duvignau, Laurent Prévot, and Y. Desalle. 2008. Toward a cognitive organization for electronic dictionaries, the case for semantic proxemy. In *Coling 2008: Proceedings of the Workshop on Cognitive Aspects of the Lexicon (COGALEX 2008)*, pp. 86–93, Manchester.

Gaume, Bruno, Fabien Mathieu, and Emmanuel Navarro. 2010. Building Real-World Complex Networks by Wandering on Random Graphs. *Information - Interaction - Intelligence*, 10(1).

[11] Language, Age, Cognitive State ...

Gu, Y. 1992. *The syntax of resultative and causative compounds in Chinese*. Ph.D. thesis, Cornell University, Ithaca, NY. unpublished.

Huang, C.R., R.Y. Chang, and S.B. Li. 2004. Sinica BOW (Bilingual Ontological Wordnet): Integration of bilingual wordNet and SUMO. In *Proceedings of the 4th LREC*, Lisbon, Portugal.

Lu, J. H.-T. 1977. Resultative verb compounds vs. directional verb compounds in mandarin. *Journal of Chinese Linguistics*, 5, 276–313.

Mei, Jia-Ju, Yi-Ming Zheng, Yun-Qi Gao, and Hung-Xiang Yin. 1984. *TongYiCi Cilin*. Commercial Press, Shangai.

Ploux, Sabine and Bernard Victorri. 1998. Construction d'espaces sémantiques à l'aide de dictionnaires informatisés. *Traitement Automatique des Langues*, 39(1), 161–182.

Resnik, Philip and Mona Diab. 2000. Measuring verb similarity. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, pp. 399–404.

Sajous, Franck, Emmanuel Navarro, Bruno Gaume, Laurent Prévot, and Yannick Chudy. 2010. Semi-automatic Endogenous Enrichment of Collaboratively Constructed Lexical Resources: Piggybacking onto Wiktionary. In H. Loftsson, E. Rögnvaldsson, and S. Helgadóttir, eds., *Proceedings of the 7th International Conference on NLP (IceTAL 2010)*, volume 6233 of *LNAI*, pp. 332–344, Reykjavik, Iceland. Springer-Verlag.

Sigman, Mariano and Guillermo A. Cecchi. 2002. Global organization of the wordnet lexicon. *Proceedings of the National Academy of Sciences*, 99(3), 1742–1747.

Thompson, Sandra A. 1973. Resultative verb compounds in Mandarin Chinese: A case for lexical rules. *Language*, 49, 361–379.

Zipf, G. K. 1949. *Human behavior and the principle of least effort*. Addison-Wesley.