Computational modeling of verb acquisition, from a monolingual to a bilingual study *

Laurent Prévot[a], Chun-Han Chang[b], and Yann Desalle[c]

[a] LPL, CNRS - Université de Provence, France
laurent.prevot@lpl-aix.fr
[b] Graduate Institute of Linguistics, National Taiwan University, Taiwan
chunhanchang@ntu.edu.tw
[c] Octogone, CNRS - Université de Toulouse, France
desalle@univ-tlse2.fr

Abstract. Going cross-linguistic is a an important but challenging track for validating a computational model of lexical organization. Our starting point is a computational model that has been established and validated on French language and we attempted to apply it on Mandarin language. The main ingredients of this model are computational lexical resources and a psycho-linguistic protocol involving extra-linguistic material (video-clips). At this stage, all the psycho-linguistic experiments have been ran, most of the resources have been built but some comparative analyses are not fully completed. Still the project is advanced enough to report on the issues we had to address while performing this cross-linguistic move concerning the resources, the analysis of the data and the data alignment across languages.

Keywords: Computational model, Acquisition, cross-linguistic study, French, Mandarin

1 Introduction

The study reported in this paper is part of a broader project (M3- Model and Measurement of Meaning) which investigates lexical organization of French and Mandarin.[1] The project combines psycho-linguistic and computational approaches for investigating the verb lexical organization of both languages. We therefore deal with two kinds of data: On the one hand the productions obtained through psycho-linguistic experiments performed with French and Mandarin Chinese speakers; On the other hand we systematically exploit existing electronic resources. A schematic organization of the project in proposed in Figure 1. Two parallel cross-linguistic studies were planned. First at the psycho-linguistic level, the same protocol, Approx (Duvignau and Gaume, 2003; Duvignau *et al.*, 2005; Duvignau *et al.*, 2007)), for analyzing the productions at different ages ($Pi.fr/Pi.tw$ = French/Taiwanese Population of Age Group i,...) would be applied to both languages for studying productions variability across languages. Then at a computational level, existing lexical resources, in our case paradigmatic graphs[2] ($PG.fr/PG.tw$ = French paradigmatic graph, Mandarin paradigmatic graph), should be aligned thanks to translation resources in order to compare the structural properties and the organization of the graphs. For example, we would like to check whether some lexical clusters in a language are preserved when projected in the other language and if it is the case which ones.

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

[1] More information about M3 project on these websites: http://erss.irit.fr/flexsem (France) http://140.112.147.149:81/m3 /(Taiwan).

[2] Paradigmatic graphs a here a sort of synonymy graph in which the synonymy relation is kept relatively loose.

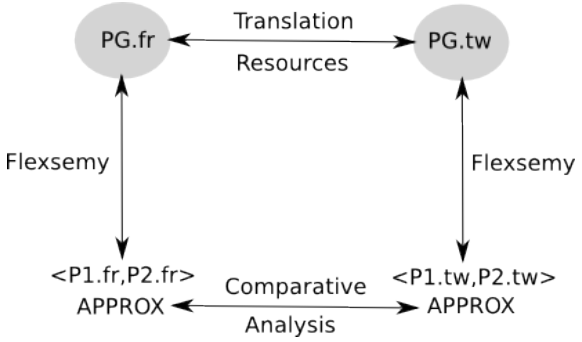

Figure 1: M3 Square

A validation across the methodologies is proposed as the following. Given a set of answers of a population, we can look at their properties in the paradigmatic graph. Then as explained in this volume (Desalle *et al.*, 2010), an hypothesis based on a measure, so-called *flexsemy* measure, ties the properties of the lexemes nodes in the graph with their psycholinguistic characteristics (which populations produced them) (see section 2). This methodological cross-validation is represented by the vertical arrows in the Figure 1.

The paper is structured as follows. We first briefly introduced the whole model build on French and sum-up some previous validation studies and experiments on French language. Then in section 3 we recap the work done to setup and experiment the model on Mandarin language. In section 4 we present the main research issues emerging from the cross-linguistic move. Sections 5 and 6 address two specific issues: the update of an analysis protocol and the data alignment. Section 7 discusses the work on a more general level and gathers the implication of going cross-linguistic for the kind computational psycholinguistics study we started from.

2 A computational model for language acquisition

The proposal is basically constituted of three ingredients: A protocol for eliciting verbal answers from subjects at various ages, a computational model of the lexicon and an hypothesis on how the typical production by age relates to the computational model.

2.1 The eliciting protocol

The material of this protocol has been created by Karine Duvignau and Bruno Gaume and consists of 17 speechless action-video clips: *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* . In each video clip, a woman alters an object from an initial state to a final state with her hands or with an instrument as illustrated in Fig 2 where the woman is about to tear up a newspaper.

Before showing the video, participants were told that “a lady” was going to perform. After showing the video, the experimenter asks the children what the woman has done. After the first answer, a rewording of the action was asked to the participant. This protocol has been used with success with adults, children of various ages as well as with patients presenting various cognitive disorderss.

The answers obtained are later analyzed according to an initial version of the Approx protocol characterizing the answers in terms of *genericity* and *conventionality*. See (Duvignau and Gaume, 2003; Duvignau *et al.*, 2005; Duvignau *et al.*, 2007) and (Desalle *et al.*, 2010) in this volume for

Figure 2: Action movie

more details on the experiment. As reported in details in (Desalle *et al.*, 2010), the results confirm that 2-4 years old children produce more approximations than adults. According to the hypothesis, these differences between young children and adults should be retrieved by the study of the lexical organization of verbs.

2.2 Computational model and hypothesis

The model is defined on a paradigmatic graph, in our case a graph built from Dicosyn,[3] a compilation of seven French synonymy dictionaries. The paradigmatic graph we work with are defined as following:

- the vertices are the lemmas
- there is an edge between two vertices if and only if at least one of the seven dictionaries states that the corresponding lexemes stand in a synonymy relation.

(Gaume *et al.*, 2002; Gaume, 2004) proved that this kind of graphs were *Small World Networks* (Watts, 1999).

Then we consider *random-walks* on these graphs. A random-walk on a graph can be though as someone wandering along the edges of the graph passing randomly from a vertex to another vertex. An important parameter of the random-walks is their length: how many steps of the wanderer do we consider? These random-walks allows to compute a *proximity* score for a vertex, a starting point[4] and a number of steps. See, (Gaume, 2004) and (Desalle *et al.*, 2010) in this volume for details on the computational model. Further, (Desalle *et al.*, 2010) has defined the *flexsemy* score based on the proximity and the degree of the vertex (how many neighbors the vertex has in the graph). It has been shown how flexsemy can be generalized to a flexsemy score of the subject and how this score predicts nearly perfectly the age group of the subject.

3 Porting the study to Mandarin language

As seen above our model relies on two main ingredients: lexical resources and productions by the subjects at various ages. This section briefly introduces and points the relevant works in which these issues have been addressed during the project.

[3] http://www.crisco.unicaen.fr/Presentation-du-dictionnaire.html

[4] In fact, one can start from any state of the graph.

3.1 Lexical Resources

The French lexical graph has been created from a compilation of synonym dictionaries. The same resource was not available for Mandarin language. Moreover the notion of lexical unit is not completely settled among the linguists. In this volume, (Gaillard *et al.*, 2010) describes how a comparable resource to our French paradigmatic graph was built for Mandarin language on the ground on existing resources such as the CiLin (Mei *et al.*, 1984) and the Chinese WordNet (Huang *et al.*, 2004).

3.2 Experiments

For the M3 project, we focused on two populations in France and in Taiwan : children without cognitive disorders and adults. We built two databases (one for each language), listing all the verbs produced by the participants for each video clip.

Taiwanese analysts faced a crucial issue with Mandarin verbal forms that are difficult to disentangle from more complex constructions. Resultative constructions, well known in Mandarin (Thompson, 1973; Lu, 1977; Gu, 1992; Cheng and Huang, 1994; Gao, 1997) are very frequent in our experimental data. In fact most of the action descriptions can be split in a description of the action itself and in a description of its result. In French this is often lexically encoded in the verb lemma, but the situation is not so clear for Mandarin. Four main cases can be considered (Li, 1990; Chang, 1998): Action verb (1-a), Resultative verb (1-b), Resultative compound (1-c), Resultative construction (1-d)-(1-e).

(1) a. 切 香菜 / cie1 xiang1tsai4 *(cut parsley)*
b. 香菜 斷-了 / xiang1tsai4 duan4-le5*(parsley snapped-PFV)*
c. 切-開 香菜 / cie1-kai1 *(cut-open parsley)*
d. 切-成-碎碎的 / cie1-chun2-sue4sue4de5 *cut-become-pieces*)
e. 切-的-碎碎的 / cie1-de5-sue4sue4-de5 *cut-DE-pieces*)

This linguistic characteristic has several consequences. First, while simple verbs are part of standard resource, only some conventionalized compounds and none of the constructions are included. This complicates the mapping of the resource-based graphs used in the computational model with the productions observed in the experiments.

Moreover, there are a large number of hapaxes in the Taiwanese data because of the resultative constructions that can be very creative like (2).

(2) 碎-的-很-可怕］/ sue4-de-hen3-ke3pa4 *(shattered-DE-very-scary)*

At this stage, Taiwanese data appeared very different from the French data and it was very unlikely that we would be able to perform precise comparison of such different data. We will see however in section 6 that a more careful look at the data lead us to revise this observation.

4 Cross-linguistics issues

The schematic view of the project Figure 1, propose an validation from a study at different perspective, and this across languages as well. Several issues have been raised however about the cross-linguistic validation.

Movies cultural bias First, in Figure 1 there is an hidden ingredient: the movies. The characterization in the Approx protocol relies on the movies themselves. As explained in (Magistry *et al.*, 2009; Cheung *et al.*, 2010) these movies may introduced a cultural bias since they are all very

familiar for French subjects but they may appear strange to Taiwanese subjects. This could lead to peculiar answers triggered by surprise or amusement. This was the main reason of the extra-linguistic study carried on the movies and that characterize each movie according to its cultural bias. The study has been carried on through an extra-linguistic questionnaire that has been answered by both French and Taiwanese subjects. This study has the effect to reject a few movies has they appeared culturally biased and therefore to reject the corresponding verbal productions as illustrated in Figure 3. See **??** in this volume for more details.

Comparability of the graphs Second, there are many possibilities for building Mandarin paradigmatic graphs. It is not straightforward to guarantee that the graph built is comparable with the French paradigmatic graph. Indeed, there was no such resource as a compilation of Mandarin synonym dictionaries. (Gaillard *et al.*, 2010) describes the methodology to build graphs that one can reasonably attempt to compare according to their lexical structure.

Data analysis As we have seen (1-c)-(1-e) the verb phrase structure is quite complex in Mandarin and it is delicate if not impossible to extract systematically a verbal form from the verb phrase. Moreover Mandarin language splits clearly the action and the result event of the actions described. This triggered the need for the analyst to have a finer grained characterization of the production of the participants. Ultimately this lead to a richer Approx protocol which is described in section 5. The new protocol is now almost complete but the data has not been fully analyzed according to this new grid of analysis.

Resource alignment Finally, while we hoped to rely mostly on existing translation resources, the resultative compound and constructions tend to not even be in monolingual dictionary, let alone bilingual ones. Moreover many verb forms were child's talk that were not necessarily present in the resource neither. Therefore we had a serious discrepancy between the experimental data and the lexicographic data. However, we want to perform a graph analysis of the experimental data as well. As a consequence, we decided to build experimental graphs that correspond to a restriction of lexicographic graphs to the forms found in the experimental data. These graph are then augmented by the forms not present in the lexicographic resources but for which a link with an existing vertex could be proposed. Finally, since we wanted to have a more precise comparison on these subgraphs, we performed a manual alignment of the French and Mandarin data as described in section 6. This was possible thanks to the relatively small size of the elicited lexicon.

5 A new protocol of analysis

In Approx, the verbal answers elicited by the action-video clips are analyzed on two levels: the responses level (general) and the verbal forms level.

5.1 Analysis of the responses

The responses are analyzed with two kinds of criteria (Table 1):

- One criterion about the linguistic productions
- Two criteria about the extra-linguistic productions:
  - Pointing gesture: did the participant point at an object in the movie?
  - Action gesture: did the participant mime the action?

Table 1: Criteria for the analysis of the response

Answer	Action gesture	Pointing gesture
Valid: focus on action	Yes	Yes
Focus on Actor	No	No
No verb in the answer		
Irrelevant answer		
No answer		
Movie no watched		
Other		

5.2 Analysis of the verbal forms: old protocol

In the previous protocol, the verbal forms that denominate a given action A were categorized in:

- Conventional verbal forms: the verbal forms belong to the semantic field of the action and fit the action A (e.g. *to peel* / peeling a banana)
- Intra-domain approximation: the verbal forms belong to the semantic field but not well fit the action (e.g. *to cut* / peeling a banana)
- Inter-domain approximation: the verbals forms 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/)

But, with these categories we could not know where the approximations come from. How to discriminate between these two inter-domain approximations for the action tearing newspapers: *to saw*, *to break*. For the first one, the approximation is on the instrument, for the second one it is on the object. Furthermore, in Mandarin, many verbal forms are compound verbs like 打-破 /da3-po4 (*hit-break*) and many of the Mandarin compound verbs in Approx are resultative compound verbs. But most of the children's approximations in Mandarin are on the resultative component of the compound verbs (破 /po4 (*break*). So, to distinguish these approximations from others, it was necessary to point out that the result expressed in these verbs doesn't fit the final state of the object transformed by the action.

5.3 Analysis of the verbal forms: new protocol

According to the previous section, we search all the possible dimensions on which the participant could produce an approximation.

As all the actions to denominate are state-changing actions, 5 dimensions where selected to analyze the verbals forms:

- Actor: The actor who perform the action in the movie;
- Entity: The entity (object) transformed by the action;
- Final State: The final state of the object;
- Instrument: The instrument (it can be a body part) used to perform the action;
- Manner: The way the action is performed.

Given a verbal form v, we analyze it on this set of dimensions D according to:

- the dependence on dimensions: which dimensions in D are meaningful for v?
- the appropriateness with the action-video clips: given a meaningful dimension and an action-video clip A, is the verbal form v appropriate to the action?

Thus, two closed questions per dimension, respectively for the dependency and the appropriateness, enable the qualification of verbs (e.g. Table 2).[5]

Table 2: An example of a closed question on Actor dimension

Criterion	Question
Actor dependency	Does the verbal form entail any Agent? a) No, no literal interpretation of this verbal form can, without tension, be performed by an Agent. b) Yes, one of the literal interpretation of this verbal form can, without tension, be performed by an Agent.
Actor appropriateness	Is the verbal form appropriate to the woman seen in the movie? a) No, the verbal form can never be appropriate to the woman. b) Yes, the verbal form can be appropriate to the woman.

Evaluation of the protocol For the time being, we have evaluated the quality of the closed questions for the Actor and Entity dimensions on a French and a Mandarin set of 40 verbs per language. Four native evaluators[6] answered the questions. The results are exposed in table 3 below:

Table 3: Results of the inter-evaluator evaluation on Actor and Entity dimensions

Criterion	**Language**	**Average agreement** [7]	**Global agreement**[8]
Actor dependency	Mandarin	96.5%	92.5%
Actor appropriateness	Mandarin	100%	100%
Entity dependency	Mandarin	100%	100%
Entity appropriateness	Mandarin	97%	95%
Actor dependency	French	98%	98%
Actor appropriateness	French	100%	100%
Entity dependency	French	100%	100%
Entity appropriateness	French	90%	85%

As inter-evaluators agreements were good on both languages (average agreement ≥ 90; global agreement ≥ 85), the questions for Actor and Entity dimensions were kept to analyze the verbal forms produced in Approx.

6 Aligning the production

Originally the project planned to use existing translation resources to propose an alignment between French and Mandarin resources. However, many productions (mostly resultative compounds and constructions) in the Mandarin data were not present in the Mandarin resources. Moreover, the translation resources available (standard bilingual dictionaries and online resources) were not fine grained enough to mention all the possible translation links between the French and the Mandarin lexical units. Since this alignment was crucial, we decided to perform a manual alignment of the productions.

[5] When the answer is *no* for the dependency question, the evaluator doesn't analyze the verbal form on appropriateness.
[6] All of them are linguists or student in linguistics.

6.1 First alignment protocol

Competence in both Mandarin and French was required to do this task. We recruited college students in Taiwan, Mandarin native speakers, who major in French or study French as second foreign language. Three raters were chosen from ten applicants based on the results of a screening task with twenty evaluation questions. The evaluation is to judge whether a pair of Mandarin and French verbs is a good translation of each other as illustrated in Table 4. In each question, there is the verb pair and the event which the verb was employed to interpret. Before the evaluation, raters were asked to watch the films, and told to think of the verbs with senses related to the event presented in the film. If the answer is positive, i.e. it is a good translation, put '1' in the evaluation column. If the answer is negative, fill in '0', as is illustrated in table. And, if the French verb can only be a good translation of certain component of the Mandarin verb, put that component in the blank. This is particularly set for Mandarin compound verbs. For example, in the compound 打-破-掉 /da3-po4-diao4 (*'hit-break-off'*), only the component 打 /da3, rather the whole compound, can be a good translation of *'donner un coup'*. In addition, the raters were reminded not to be too strict with the translation since participants of the original experiment could be as young as 3-year-old.

Film	Mandarin	French	Evaluation
break glass	打-破-掉 /da3-po4-diao4	donner un coup	打 /da3
crush tomato	掉 /ya1	presser	1
peel orange	掉 /buo1	retirer	0

Table 4: Alignment protocol

We evaluated the translation on a subset that has been aligned by 3 students and a pair of experts. The κ-score ranges from 0.35 to 0.59 which is far from perfect. However, this was partly explained by the level in French among the students who was not the same and they got a different idea of not being too strict in the alignment. We therefore needed to perform a final check on this data.

We noted while performing these alignments that Mandarin had a lot of variety in the result description while French due to the post-treatment of the data, that reduced the answers to a verb lemma was much poorer. We therefore went back in the original French transcripts and find out that in fact in some cases, relevant constructions quite comparable with Mandarin construction had been discarded when selecting the verbal form (3).

(3) a. elle a découpé le pain en deux morceaux *(she cut the bread into two pieces)* $\rightsquigarrow$ découper
b. elle a partagé en deux une flute de pain *(she cut into two a bread)* $\rightsquigarrow$ partager
c. elle coupe en morceaux les brocolis *(she cut into pieces the broccolis)* $\rightsquigarrow$ couper

On the other hand, the Taiwanese data featured a lot of hapaxes like (2) that we removed for performing a second alignment. Some hapaxes differed from each other only by little like (4). For this category, we tried to keep the common elements of both productions if they could be gathered in homogeneous group easily. After close analysis, French and Mandarin data were not that different in the end. Only the respective frequency of the phenomena differs.We therefore performed a second alignment, for the new French forms and the reduced number of Mandarin constructions. About the later, we kept the non-hapaxes and try to keep a generic form in case several hapaxes share the same material.

(4) a. 鋸-成-兩半 / ju4-chun2-liang3ban4 *(saw-become-two-halves)*

b. 鋸-成-兩截 / ju4-chun2-liang3jie2 *(saw-become-two-sections)*
c. 鋸-成-兩段 / ju4-chun2-liang3duan4 *(saw-become-two-sections)*
d. 鋸-成-兩片 / ju4-chun2-liang3pian4 *(saw-become-two-slices)*

6.2 Additional alignments

For the second evaluation, a few amendments were made to the original protocol. Information about the film was omitted. The raters should already have an idea about the events after the first evaluation task. In case of doubt, the raters were asked to fill in '1'. They were also reminded not to overuse the component option, explained in previous section and again not to be too strict with the translation, for the Mandarin verbs can be productions of children as young as 3-year-old. Here again we evaluated the alignment against a subset of pairs aligned by experts. The κ-score is 0.56 which is reasonable but still calls for improving the alignment protocol.

7 Conclusion and Future work

While the original model seems robust and applicable to all languages, its actual implementation on a typologically remote language required quite a little bit on adaptation and careful thinking. More interestingly, applying the model to Mandarin lead us to refine by many aspects the original analysis of the French data. The figure 3 sums-up these updates and should be compared to the original design of the study Figure 1.

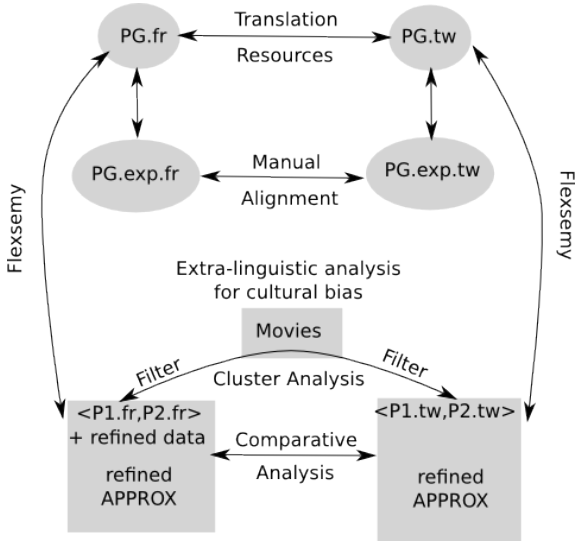

Figure 3: M3 Square updated

Compared to the original plan, we had to introduce an intermediate resource, the experimental graphs that are a sort of compromise for studying lexical graph structures across languages in a precise way. While the psycholinguistic analysis for both languages has been done, the study made clear that it was necessary to refine the Approx protocol. Finally, the movies were thought to be cultural-independent and this property was used in deriving conclusions. It became clear in the project that systematic deeper assessments about cultural-independence on extra-linguistic material was an important parameter to bring in for analyzing the results and drawing conclusions.

As for future work, the original plan is almost completed. The main part that remains to be done by the end of the project is the structural comparison of the two pairs of graphs. The

mathematical tools for this comparison are now ready (Gaillard *et al.*, 2010).

The new design of the Approx has been particularly time consuming. It is ready for two dimensions of analysis on which we have shown satisfactory agreement results. The other dimensions are ready as well but have not been evaluated yet. Also, it is only in the future, when people will use this revised protocol for French, Mandarin or another language that we will get the full benefit of the work on this part.

Taking into consideration verbal constructions (e. g. *couper en deux / cut into two*) is important for cross-linguistic comparison but the integration of these constructions into a computational model grounded on paradigmatic graphs is not straightforward. The easiest way is to treat the most conventionalized constructions as locutions and therefore as lexically units entitled to appear in the paradigmatic graph. This is also suggested by the fact that a few of those constructions actually are in existing resources. This is the case of *'couper en morceaux' / cut into pieces* which is present in Dicosyn. However, deciding which constructions should be included and which one should not will not be easy to decide, in particular for Mandarin in which these constructions are very productive and present a variety of conventionalization levels.

This may suggest the necessity to go further on the issue of compositionality, also bringing in syntactic considerations, for this family of computational models or at least how to use them in a compositional study. This line of investigation is related to the recent works that are attempting to make sense of Distributional Semantic Models with compositionality (Lenci, 2008; Baroni and Zamparelli, 2010; Magistry *et al.*, 2010).

References

Baroni, M. and R. Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

Chang, Claire Hsun-Huei. 1998. V-V compounds in mandarin chinese: Argument structure and semantics. In Jerome L. Packard, ed., *New approaches to Chinese word formation: Morphology, phonology and the lexicon in modern and ancient Chinese*, pp. 77--101. New York: Mouton de Gruyter.

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.

Cheung, Hintat, Yann Desalle, Karine Duvignau, Bruno Gaume, Chunhan Chang, and Pierre Magistry. 2010. The use of a cultural protocol for quantifying cultural variations in verb semantic between chinese and french. In *Proceedings of PACLIC 24*, Sendai, Japan.

Desalle, Yann, Bruno Gaume, Shu-Kai Hsieh, and Hintat Cheung. 2010. Towards an automatic measurement of verbal lexicon acquisition: the case for a young children-vs-adults classification in French and Mandarin. In *Proceedings of PACLIC 24*, Sendai, Japan.

Duvignau, Karine, M. Fossard, Bruno Gaume, M.-A. Pimenta, and J. Elie. 2007. Semantic approximations and flexibility in the dynamic construction and deconstruction of meaning. *Linguagem em Discurso*, 7(3), 371--389., Sept-Dec.

Duvignau, Karine and Bruno Gaume. 2003. Linguistic, psycholinguistic and computational approaches to the lexicon: Contributions to early verb-learning. *Journal of the European Society for the Study of Cognitive Systems*, 6(1).

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*.

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 *Proceedings of PACLIC 24*, Sendai, Japan.

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

Gaume, Bruno. 2004. Balades aléatoire dans les petits mondes lexicaux. *Information Engineering Sciences*, 4(2).

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.

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

Huang, Chu-Ren, Ru-Yng Chang, and Hsiang-Pin Lee. 2004. Sinica BOW (bilingual ontological wordnet): Integration of bilingual WordNet and SUMO. In *Proceedings of LREC*, Portugal.

Lenci, A. 2008. Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, 20(1), 1--31.

Li, Yafei. 1990. On V-V compounds in chinese. *Natural Language and Linguistic Theory*, 8, 177--207.

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

Magistry, Pierre, Laurent Prévot, Hintat Cheung, Chien yun Shiao, Yann Desalle, and Bruno Gaume. 2009. Using extra-linguistic material for mandarin-french verbal constructions comparison. In *Proceedings of PACLIC 23*, HongKong.

Magistry, Pierre, Laurent Prévot, and ShuKai Hsieh. 2010. Classifier-noun composition in mandarin: A testcase for distributional semantic models. In *Proceedings of Compositionality and Distributional Semantic Models workshop (ESSLLI)*.

Mei, Jia-Ju, Yi-Ming Zheng, Yun-Qi Gao, and Hung-Xian Yin. 1984. *TongYiCi CiLin*. Commercial Press, Shanghai.

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

Watts, D.J. 1999. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press.