Skillex, an action labelling efficiency score: the case for French and Mandarin

Yann Desalle[1], **Bruno Gaume**[1], **Karine Duvignau**[1], **Hintat Cheung**[2], **Shu-Kai Hsieh**[3], **Pierre Magistry**[3], **Jean-Luc Nespoulous**[4]
{yann.desalle, gaume, duvignau}@univ-tlse2.fr, hintat@ied.edu.hk, shukaihsieh@ntu.edu.tw, pmagistry@gmail.com, nespoulo@univ-tlse2.fr
[1] CLLE-ERSS, Université de Toulouse, France
[2] The Hong Kong Institute of Education, China
[3] Graduate Institute of Linguistics, National Taiwan University, Taiwan
[4] Octogone, Université de Toulouse, France

Abstract

We propose a model to compute two measurements of semantic efficiency of verbs as action labels. It is based on the exploration of the specific structure of synonymy networks of verbs. We use these measurements to analyse and compare the semantic efficiency of [Children/Adults] productions in action labelling tasks, in French and Mandarin. The combination of these two measurements leads to a generic score of semantic efficiency, *Skillex*. Assigned to participants of the *Approx* protocol experiment, this score enables us to accurately classify them into Children and Adults categories, be they French or Mandarin native speakers.

Keywords: lexical acquisition, lexical networks, action labelling.

Introduction

Most research on lexical acquisition of action verbs shows that, in order to label actions, young children produce action verbs that are semantically less efficient than adults'ones.

In this article, we tackle this issue in a different way. Instead of using psycholinguistic criteria (specificity, conventionality, imageability ...) to evaluate how efficient a verb is to label an action , we automatically compute the semantic efficiency of verbs by mapping the semantics of labelled actions onto a synonymy network of verbs and by exploring the specific structure of such a lexical network.

Thus, we propose a model to compute a generic score of semantic efficiency, called *Skillex*, which combines two other efficiency measurements, *Prox* and *Deg*. Assigned to participants of the *Approx* protocol experiment, *Skillex* accurately categorizes them into the two [Young Children/Adults] age groups, be they French or Mandarin native speakers.

First, state of the art research about semantic acquisition of action verbs and current work on lexical networks are briefly reviewed. Then we detail our model and its evaluation, on the basis of the *Approx* protocol. The last section is devoted to the conclusion.

Semantic Acquisition of action verbs

From works of Bowerman (1974), Schaefer (1979) and Gentner (1982) to recent research (Bowerman, 2005; Hirsh-Pasek & Golinkoff, 2006; Duvignau, Fossard, Gaume, Pimenta, & Elie, 2007; Imai et al., 2008), action verbs acquisition dynamics have consistently been shown to be distinct from object nouns acquisition dynamics. This difference can be explained by the genuinely relational nature of verb predicates, which strongly constrains the categories of the units they put in relation. Since, in addition, many verb predicates induce categories that are not natural (contrary to most object nouns), verb acquisition is more difficult, thus slower, than noun acquisition[1].

Recent research (Bowerman, 2005; Duvignau et al., 2007; Gaume, Duvignau, Prévot, & Desalle, 2008) has discovered two salient patterns in the verb productions of young children: (a) verbs that, although semantically close to the expected conventional verb, don't match the labelled action on at least one of their semantic components (b) verbs that expect generic categories on their semantic components: many objects fit in such categories.

Lexical networks

A graph[2] is defined by a set of vertices V and a symmetric relation $E \subseteq \mathbf{P}_2^V$ (the set of 2-element sets of V).When V is a set of lexical units and E a lexical relation (e.g. synonymy, co-occurrence...), the graph $G = (V, E)$ is called a lexical network. The past decade has seen several works (Gaume, 2004; Gaume et al., 2008; Steyvers & Tenenbaum, 2005; De Deyne & Storms, 2008a; Morais, Olsson, & Schooler, 2013) that show that most lexical networks, as most terrain networks[3], are Hierarchical Small World Networks (HSWN), insofar as they share similar properties (Watts & Strogatz, 1998; Barabasi & Réka, 1999; Newman, 2003; Gaume, Mathieu, & Navarro, 2010):

- $\mathbf{p_1}$: **Low density** (not many edges)
- $\mathbf{p_2}$: **Short paths** (the average number L of edges of the shortest path between two vertices is low)
- $\mathbf{p_3}$: High **clustering rate** C (locally densely connected subgraphs can be found whereas the whole graph is globally sparse in edges)
- $\mathbf{p_4}$: Degree Distribution akin to a **power law**[4].

In this paper we focus on 2 synonymy graphs :

- **A French dictionary graph of verbs:** $G_F = (V_F, E_F)$ is a graph of the verbs extracted from DicoSyn[5]: there is an

[1] On this issue, note also the perspectives of Tardif (1996); Choi and Gopnik (1995).

[2] Here we only consider undirected graphs.

[3] Terrain networks are networks that model real data, for example in sociology, linguistics or biology. They contrast with artificial networks (deterministic or random).

[4] Although Morais et al. (2013) showed that the distribution is more correlated with a "power law with exponential cut-off" than with a power law.

[5] DicoSyn is a compilation of synonymy relations extracted from seven other dictionaries (Bailly, Benac, Du Chazaud, Guizot, Lafaye, Larousse and Robert). DicoSyn was first

edge $\{A, B\}$ if the verbs described by the vertices A and B are synonyms in DicoSyn. G_F is made symmetric and reflexive.

- **A Mandarin dictionary graph of verbs:** $G_M = (V_M, E_M)$ is a graph of verbs extracted from CilinCWN: a fusion of Chinese WordNet (CWN)[6] and a Chinese thesaurus TongYiCi CiLin (Cilin)[7]. Data was processed similarly to the way G_F was built.

Table 1 shows the typical HSWN properties of G_F and G_M on their largest connected component. Our approach consists in exploiting the HSWN properties of these synonymy networks. The two properties p_3 and p_4 are especially useful.

Table 1: Properties of G_F and G_M on the largest connected component : n: number of vertices, m: number of edges, L: average path length, C: clustering coefficient, $\lambda(r^2)$: slope of the power-law model (in brackets: correlation coefficient with the model).

	n	m	L	C	$\lambda(\mathbf{r^2})$
$\mathbf{G_F}$	8993	111659	4.19	0.14	-2.02 (0.93)
$\mathbf{G_M}$	8386	94198	5.68	0.61	-2.30 (0.86)

Several research projects have demonstrated a relation between the structure of lexical networks and the lexical acquisition process. According to Steyvers and Tenenbaum (2005), in lexical networks built from the Roget's thesaurus, WordNet and the USF word association norms (Nelson, McEvoy, & Schreiber, 2004), vertex degrees are correlated with:

- the age of acquisition (AoA) of English words (Morrison, Chappell, & Ellis, 1997)
- the frequency of occurrence of such words in English, itself correlated with their AoA (Kučera, Francis, Carroll, & Twaddell, 1967).

These findings are confirmed by a study of De Deyne and Storms (2008a), for the Dutch language, on the basis of the graph extracted from the Dutch Word Association norms (De Deyne & Storms, 2008b). The study also shows that both the clustering coefficient in the word's neighbourhood (distance 2) and its betweenness centrality (measure of the centrality of a vertex in a graph) are correlated to its AoA.

Model

Theoritical motivations of the model

Our model is motivated by the parallel between (a) experimental results on semantic acquisition of action verbs and (b) our hypotheses on HSWN properties of synonymy networks (Duvignau, Gaume, & Nespoulous, 2004; Gaume et al., 2008):

- **1.a** Verbs produced by adults are more specific than those produced by children
- **1.b** Specific verbs' degrees are low ($\mathbf{p_4}$)
- **2.a** Action verbs produced by children are less appropriate to the labelled actions than those produced by adults
- **2.b** In synonymy networks, verbs are brought closer if their meanings are closely related ($\mathbf{p_3}$).

This model is based on two measures : (1) the degree of a verb in a synonymy network and (2) a verb's proximity to a lexico-semantic zone of a synonymy network. In a verb synonymy graph $G = (V, E)$, the degree of a verb $v \in V$, denoted by $deg(v)$[8] is its number of neighbours in the graph. A verb's proximity to a lexico-semantic zone, however, is a more complex measure, and will be detailed in the next section.

Prox

Consider $G = (V, E)$, a verb synonymy graph with $n = |V|$ vertices and $m = |E|$ edges, such that G is reflexive (any vertex is linked to itself). Consider a traveller randomly walking along edges of the graph G, from vertex to vertex:

- At each moment $t \in \mathbb{N}$ the traveller is on a vertex $u_t \in V$
- At time $t + 1$, the traveller reaches one neighbour of u_t, randomly chosen with uniform probability.

This process is called a *simple random walk* on G, as described for example by (Bollobas, 2002). It is formally described by a *Markov Chain* on V, with a $n \times n$ transition probability matrix $[G]$:

$$[G] = (g_{u,v})_{u,v \in V}, \text{ with } g_{u,v} = \begin{cases} \dfrac{1}{\deg(u)} \text{ if } \{u, v\} \in E, \\ 0 \text{ else.} \end{cases} \quad (1)$$

Since G is reflexive, no vertex has a null degree, $[G]$ is therefore definite. By construction, $[G]$ is stochastic: $\forall u \in V, \sum_{v \in V} g_{u,v} = 1$.

Let us define a lexico-semantic zone of the graph G by a probability distribution Δ on V, its vertex set (more details on such a definition are given hereafter). After a given number of steps $t \in \mathbb{N}$, a random walker who started its walk from the Δ distribution, at a time $t_0 = 0$, has a probabilistic location on V, described by the probability distribution $\Delta[G]^t$. When the starting vertex (u) is known, the Δ_u starting distribution is actually the probability 1 to be on vertex u. In that case, the probability for the walker to be on a vertex v after t steps is $(\delta_u[G]^t)_v = [G]^t_{u,v}$. As in (Gaume, 2004), the Perron-Frobenius theorem (Stewart, 1994) then helps us show that:

$$\forall u, v \in V, \lim_{t \to \infty} (\delta_u[G]^t)_v = \lim_{t \to \infty} [G]^t_{u,v} = \frac{\deg(v)}{\sum_{x \in V} \deg(x)} \quad (2)$$

This means that, as t grows to infinity, the probability of being on a vertex v at time t does not depend on the starting vertex, it becomes simply proportional to v's degree. However, in the early stages of the walk, the probability distribution heavily depends on u, the starting vertex. When t is

compiled at ATILF (Analyse et Traitement Informatique de la Langue Française), before being corrected at CRISCO laboratory (http://elsap1.unicaen.fr/dicosyn.html) (Ploux & Victorri, 1998).

[6] 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, Chang, & Lee, 2004), at the time of writing it contains 28,815 synonyms.

[7] The Tongyici Cilin (Mei, Zheng, Gao, & Yin, 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.

[8] $deg(v) = |\{u \mid \{v, u\} \in E\}|$.

small, the vertices most probably reached by the walker are vertices that are *close* to u, insofar as many short paths link them to u[9]. So, for a small t, the probability distribution is a good indication of the closeness of two vertices in the network. In the following, we thus consider short random walks with the fixed parameter $t = 4$[10]. We then define the proximity of a verb $v \in V$ to a lexico-semantic zone defined by a probability distribution Δ by:

$$prox(v,\Delta) = \frac{(\Delta[G]^4)_v}{\max_{r \in V}(\Delta[G]^4)_r} \tag{3}$$

For example, Table 2 provides the list of the 20 closest French verbs[11] to *écorcer* (*to bark*) in G_F. ($\Delta = \delta_{ecorcer}$, the certainty to be located on *écorcer*).

Table 2: The 20 closest verbs to *écorcer* in G_F, (with $t = 4$).

1) écorcer* (put the bark off)	11) écorcher (skin)
2) dépouiller* (strip)	12) écaler (husk)
3) peler* (peel)	13) voler (steal)
4) tondre* (mow, shear)	14) tailler (prune)
5) ôter (remove)	15) râper (grate)
6) éplucher (peel, pare)	16) plumer (pluck)
7) raser (shave)	17) gratter (scrape)
8) démunir (divest)	18) enlever (remove)
9) décortiquer* (steal)	19) désosser (bone)
10) égorger (slit the throat of)	20) déposséder (dispossess)

Efficiency of a verb

Let $G = (V,E)$ be a verb synonymy graph, $v \in V$ a verb and Δ_a the probability distribution on V that delimits the meaning of an action a. We define the *efficiency* of verb v in relation to Δ_a by:

$$s(v,\Delta_a) = \frac{prox(v,\Delta_a)}{deg(v)} \tag{4}$$

Our model is based on the hypothesis that adults produce verbs that have a better efficiency in relation to Δ_a than the efficiency of verbs produced by children to label the same action. The measures $prox(v,\Delta_a)$ and $deg(v)$ both play a meaningful part in the efficiency in relation to the Δ_a score :

- $prox(v,\Delta_a)$: the greater the proximity of verb v to Δ_a, the more semantically appropriate the verb v is, to describe a
- $deg(v)$: the smaller the degree of verb v, the more specific the verb v.

Four scores

This section details how our model attributes four scores of lexical performance to each individual, given a language L, a graph $G_L = (V_L,E_L)$, and a set of actions $A = \{a_1,\cdots,a_i,\cdots a_n\}$. Let $\Delta^L = \{\Delta^L_{a_1}\cdots,\Delta^L_{a_i}\cdots,\Delta^L a_n\}$ be the lexico-semantic zones that correspond, in G_L, to the actions of A. Let x be an individual who produced a set of verbs $W_{a_i,x}$ to label action a_i.

For each verb set $W_{a_i,x}$ such that $W_{a_i,x} \cap V_L \neq \varnothing$, the following figures are computed:

- $D(W_{a_i,x})$ is the mean[12] of the set $\{deg(v) \mid v \in W_{a_i,x} \cap V_L\}$
- $P(W_{a_i,x})$ is the mean of the set $\{prox(v,\Delta^L_{a_i}) \mid v \in W_{a_i,x} \cap V_L\}$
- $S(W_{a_i,x})$ is the mean of the set $\{s(v,\Delta^L_{a_i}) \mid v \in W_{a_i,x} \cap V_L\}$.

These three figures are the basis on which we compute the four scores of each participant x for the action category (e.g. motion actions, breaking/cutting actions...) defined by A:

- **Productiveness score** $N_A(x)$**:** mean of $\{|W_{a,x}| \mid a \in A\}$
- **Degree score** $D_A(x)$**:** mean of $\{D(W_{a,x}) \mid a \in A \text{ and } W_a \cap V_L \neq \varnothing\}$
- **Prox score** $P_A(x)$**:** mean of $\{P(W_{a,x}) \mid a \in A \text{ and } W_a \cap V_L \neq \varnothing\}$
- **Skillex score** $S_A(x)$**:** mean[13] of $\{S(W_{a,x}) \mid a \in A\}$.

Evaluation

Approx protocol

The *Approx* protocol (Méligne et al., 2011; Duvignau, Tran, & Manchon, 2013) is, on average, completed by a participant in 20 minutes, and enables us to compute a lexical performance score for each participant.

Material and Participants The material consists in seventeen 30-second action-films without speech, that show acts of separation/deterioration of objects. In each film[14], a woman alters an object with the help of her hands or with an instrument, explicitly showing an initial state and a final state .

In this paper we focus on 4 groups of participants, French and Mandarin native speakers[15] :

- C_F : 74 French young children (2-5 years old)
- A_F : 76 French young adults (18-40 years old)
- C_M: 29 Mandarin young children (2-5 years old.
- A_M: 60 Mandarin young adults (18-30 years old)

Procedure : The films are randomly shown to a participant. After each film, the experimenter asks the participant what the woman did. Between each action film, a distractor is shown to avoid perseveration effects. Results of participants who do not watch all 17 films are not taken into account. Lexical action labels are extracted from the elicited responses, and lemmatized. Compound labels are split according to their components :

- 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 (Li & Thompson, 1981)(it is not useful in French)

[9]They belong to the same cluster as u.

[10]The average path length (L) of G_F is $L = 4.19$, and G_M's is $L = 5.68$ (Table (1)). So, with a walk of length 4, a walker starting from any vertex can reach most vertices of the graph, whether we consider G_F and G_M.

[11]Verbs with an asterisk are neighbours of *écorcer*.

[12]Mean is for arithmetic mean.

[13]When $W_{a,x} \cap V_L = \varnothing$ we assign $S(W_{a,x}) = 0$.

[14]Burst a balloon, Crumple a piece of paper, Break a glass, Crush a tomato, Tear a newspaper off, Peel a carrot, Peel an orange, Put the bark of a log off, Undress a doll, Take legos down, Peel a banana, Make bread-crumb, Cut a bread, Break off a piece of bread, Chop parsley, Saw a plank of wood, Remove a sleeve.

[15]Participants don't have any cognitive issue, native speakers of Mandarin completed the protocol in Taiwan, native speakers of French, in France.

This procedure to record action labels both in French and Mandarin enables us to compare the French and Mandarin analyses reported hereafter.

From action-stimuli to lexico-semantic zones

In a graph ($G_F = (V_F, E_F)$ for French and $G_M = (V_M, E_M)$ for Mandarin), a lexico-semantic zone is the distribution of probability that denotes, as objectively as possible, an action-stimulus of the protocol. To define this distribution (in French), a mixed[16] Pop_F sample of participants is gathered by randomly choosing 25 participants from C_F and 25 from A_F. For each action a, each verb v of V_F is attributed the frequency $freq_a^F(v)$ with which it was used by participants of Pop_F to label action a.

The probability distribution Δ_a^F, on V_F, then defines a's lexico-semantic zone in G_F:

$$\forall v \in V_F, (\Delta_a^F)_v = \frac{freq_a^F(v)}{\sum_{s \in V_F}, freq_a^F(s)} \qquad (5)$$

Similarly, Pop_M, $freq_a^M(v)$ and Δ_a^M are defined for Mandarin in relation to G_M.

Task 1: Computing participant's scores

We refer to the 17 action-stimuli of the protocol as $A = \{a_1, \cdots, a_i, \cdots, a_{17}\}$, and to their corresponding lexico-semantic zones as $\Delta^F = \{\Delta_{a_1}^F \cdots, \Delta_{a_i}^F \cdots, \Delta_{a_{17}}^F\}$. Three scores $D_A(x)$, $P_A(x)$ et $S_A(x)$ are computed for each native French[17] speaker participant to the *Approx* protocol on the action category "separation/deterioration of objects" denoted by A.

In order to evaluate our model on the basis of this task, we compare young children's scores to scores of adult participants: a significant difference would mean that such scores accurately discriminate the two age groups [Children/Adults].

Task 2 : Automatic [Children/Adults] age group categorization

This task is detailed using the French case as a generic example, the exact same procedure is done for Mandarin. It consists in measuring the accuracy of the automatic categorization of the two age groups C_F and A_F, on the basis of the three scores computed in Task 1. With each of the 3 scores, we use the k-means algorithm ($k = 2$) (Hartigan & Wong, 1979) to separate the set of participants into two categories. When considering the *Degree score*, the category with the greatest centroid is assigned to the *young children* category, the other to the *adults* category. Conversely, when considering the *Prox score* or the *Skillex score*, the category with the greatest centroid is assigned to the *adults* category, the other to the *young children* category.

The accuracy of the automatic categorization is measured by the agreement rate between the expected categories (C_F and A_F) and the score-computed categories.

[16]So that lexico-semantic zones do not induce a bias towards the adult or child age group.

[17]The same three scores are computed for Mandarin speakers on the basis of $\Delta^M = \{\Delta_{a_1}^M \cdots, \Delta_{a_i}^M \cdots, \Delta_{a_{17}}^M\}$ in G_M.

Results

Task 1 results We used an ANOVA to measure how significant the difference between young children's and adults' PROX scores is, and a non-parametric Man-Whitney-Wilcoxon test to measure how significant the differences were between the *Productiveness*, *Degree* and *Skillex* scores of young children and adults[18].

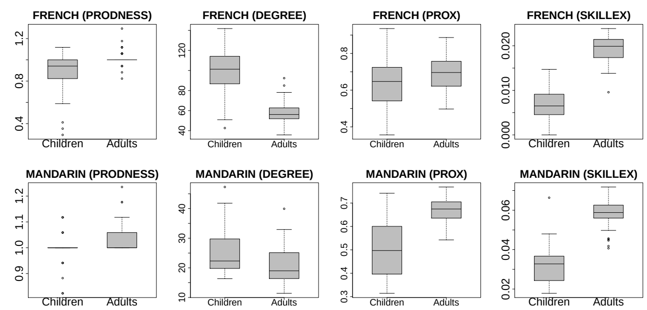

Figure 1: *Productiveness*, *Degree*, *Prox*, *and Skillex* scores of the [Children/Adults] age groups for French and Mandarin: box-and-whisker diagrams.

Results show:

- A significant difference between the *productiveness scores* of children and adults, both in French ($W(150) = 4788$; $p \approx 0$) and Mandarin ($W(89) = 1046$; $p < .05$)
- A significant difference between the *degree scores* of children and adults, both in French ($W(150) = 5389$; $p \approx 0$) and Mandarin ($W(89) = 1243$; $p < .001$)
- A significant difference between the *Prox scores* of children and adults, both in French ($F(150) = 13.58$; $p < .001$) and Mandarin ($F(89) = 79.41$; $p \approx 0$)
- A significant difference between the *Skillex scores* of children and adults, both in French ($W(150) = 5756$; $p \approx 0$) and Mandarin ($W(89) = 1670$; $p \approx 0$)

The *Productiveness*, *Degree*, *Prox* and *Skillex scores* highlight a significant difference between the verb productions of young children and of adults, upon a task that consists in labelling actions that show deteriorations or separations of objects, in both the Mandarin and French languages.

Task 2 results Task 2 aims to confirm that Task 1 results are significant and consistent enough to enable automatic categorization of adults and children. It is evaluated by the rate of agreement between automatically computed categories and expected categories, which is measured with the Precision and the Kappa of Cohen κ (Cohen, 1960):

- Precision is the observed agreement probability p_o
- The κ is defined as : $\kappa = \frac{p_o - p_e}{1 - p_e}$ in which p_e is the expected agreement probability knowing (a) the distribution of individuals on the Adult and Child categories that were built by the $2 - mean$ algorithm and (b) the distribution of individuals on the expected C_F and A_F groups.

Table 3: 2-means clustering results for French and Mandarin: *Productiveness*, *Degree*, *Prox* and *Skillex* scores

LANGUAGE		SCORE			
		$Prod^{ness}$	*Degree*	*Prox*	*Skillex*
		n=89			
Mandarin	Precision	.44	.64	.84	**.91**
	κ	.03	.17	.62	**.80**
		n=150			
French	Precision	.71	.91	.61	**.96**
	κ	.42	.81	.21	**.92**

Table 3 suggests that the main component (*degree* or *prox*) of the lack of efficiency in action labelling during lexical acquisition depends on the language to acquire: (a) whereas, in Mandarin, the *Prox score* categorizes [Children/Adults] with a substantial agreement (according to the scale of Landis and Koch (1977), $\kappa = .62$), this is not the case in French ($\kappa = .21$); (b) whereas, in French, the *Degree score* categorizes [Children/Adults] with an almost perfect agreement ($\kappa = .81$), this is not the case in Mandarin ($\kappa = .17$).

In fact, the *Skillex score* is the only score able to highlight differences of semantic efficiency of action labelling between children and adults independently from the language. It is the only score that accurately categorizes participants into the two [Children/Adults] age groups in both languages (almost perfect agreement). Furthermore, the *Skillex score*'s agreement rates are better than these of (a) in French, the degree score *D* (Precision:+5%; κ:+14%) and (b) in Mandarin, the proxemy score *P* (Precision:+8%; κ:+29%).

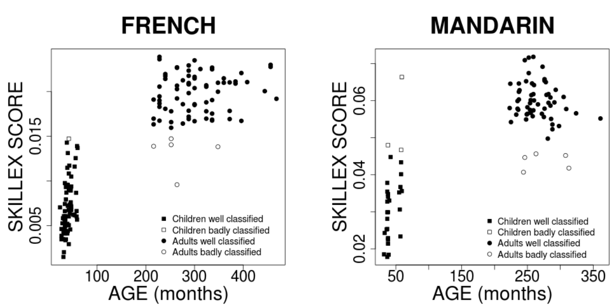

Figure 2: *Skillex score* by age group: French and Mandarin

Conclusion

The *Approx* protocol is directly applicable to different languages and participant samples (adults, children, participants with pathologies ...). It enables researchers to gather meaningful linguistic data that can be used to comparatively analyse languages and participant types.

Moreover, although this study only focuses on the two dictionary-based graphs G_F et G_M, results do not significantly vary with the dictionary on which score computations are based. In fact, despite significant local variations that reflect the differences between dictionaries, the overall structure of such lexical networks does not significantly vary. They exhibit similar degree distributions and clusters, which are common to most synonymy dictionaries of a given language, as seen in (Gaillard, Gaume, & Navarro, 2011). The participant's *Skillex score* computation, which relies on the hierarchical distribution of degrees and on the small world structure of synonymy dictionaries, is therefore robust to resource variation.

We intend to further this initial study into three directions: (a) to extend the analysis to other languages, with the long term perspective of initiating a language typology of lexical acquisition dynamics (i.e. with multilingual ressources like Wiktionary[19]), (b) to extend the protocol to other action categories (for example verbs of movement) in order to compare lexical acquisition dynamics across action types, and (c) to extend the study to the analysis of pathologies:

Various stages of the Alzheimer's disease (Joubert et al., n.d.). Building on works of Méligne et al. (2011) we formulate the two following hypotheses : On the basis of their *Approx* protocol verb production, participants can be attributed a *Skillex score* that:

- ($H1.1$) will accurately categorize participants into two [Moderate Alzheimer/Older without pathology] groups
- ($H1.2$) will NOT enable their accurate categorization into two [Moderate Alzheimer/Child without pathology] groups.

Asperger's syndrome(Atwood, 1998). Building on works of Maffre et al. (2012) we formulate the two following hypotheses : On the basis of their *Approx* protocol verb production, participants can be attributed a *Skillex score* that:

- ($H2.1$) will accurately categorize participants into two [Asperger child/Child without pathology] groups
- ($H2.2$) will NOT enable their accurate categorization into two [Asperger Child/Adult without pathology] groups.

References

Atwood, T. (1998). *Asperger's syndrome*. Jessica Kingsley Publishers.

Barabasi, A.-L., & Réka, A. (1999, October). Emergence of scaling in random networks. *Science*, *286*, 509-512.

Bollobas, B. (2002). *Modern Graph Theory*. Springer-Verlag New York Inc.

Bowerman, M. (1974). Learning the structure of causative verbs: a study in the relationship of cognitive, semantic, and syntactic development. *Papers and Reports on Child Language Development*, *8*, 142-178.

Bowerman, M. (2005). Why can't you 'open' a nut or 'break' a cooked noodle? learning covert object categories in action word meanings. In L. Gershkoff & D. Rakison (Eds.), *Building object categories in action word meanings* (p. 209-244). Mahwah, NJ: Lawrence Erlbaum Associates.

Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: a cross-linguistic study. *Journal of Child Language*, *22*, 497-529.

[18] Since, according to the Shapiro-Wilk test, the distribution of the *Productiveness*, *Degree* and *Skillex* scores are not normal distributions, ANOVA was not applicable.

[19] http://www.wiktionary.org/

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*(20), 27-46.

De Deyne, S., & Storms, G. (2008a). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*(1), 213-231.

De Deyne, S., & Storms, G. (2008b). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, *40*(1), 198-205.

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

Duvignau, K., Gaume, B., & Nespoulous, J.-L. (2004). Proximité sémantique et stratégies palliatives chez le jeune enfant et l'aphasique. In J.-L. Nespoulous & J. Virbel (Eds.), *Handicap langagier et recherches cognitives : apports mutuels* (Vol. 31-32, p. 219-255). Belgique: Revue Parole.

Duvignau, K., Tran, T., & Manchon, M. (2013). For a new look at 'lexical errors': Evidences from semantic approximations with verbs in aphasia. *Journal of psycholinguistics Research*, *42*(4), 339-347.

Gaillard, B., Gaume, B., & Navarro, E. (2011). Invariants and variability of synonymy networks: Self mediated agreement by confluence. In *Proceedings of the TextGraphs-6 2011 : Graph-based Methods for Natural Language Processing*. Portland, Oregon, USA.

Gaume, B. (2004). Balades aléatoires dans les petits mondes lexicaux. *Information Engineering Sciences*, *4*(2), 39-96.

Gaume, B., Duvignau, K., Prévot, L., & Desalle, Y. (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)* (p. 86-93). Manchester.

Gaume, B., Mathieu, F., & Navarro, E. (2010, september). Building real-world complex networks by wandering on random graphs. *I3 Information Interaction Intelligence*, *10*(1), 73-91.

Gentner, D. (1982). Why Nouns are Learned Before Verbs:. In S. A. Kuczaj (Ed.), *Language development: Language, thought and culture* (Vol. 2, p. 301-334). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hartigan, J., & Wong, M. (1979). A K-means clustering algorithm. *Applied Statistics*, *28*, 100-108.

Hirsh-Pasek, K., & Golinkoff, R. M. (Eds.). (2006). *Action meets word: How children learn verbs*. New York: Oxford University Press.

Huang, C.-R., Chang, R.-Y., & Lee, S.-B. (2004, May). Sinica BOW (bilingual ontological WordNet): Integration of bilingual WordNet and SUMO. In *4th International Conference on Language Resources and Evaluation.*

Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M., et al. (2008). Crosslinguistic studies on novel noun and verb learning. *Child Development*, *79*, 979-1000.

Joubert, S., Brambati, S., Ansado, J., Barbeau, E., Felician, O., Didic, M., et al. (n.d.). The cognitive and neural expression of semantic memory impairment in mild cognitive impairment and early alzheimer's disease. *Neuropsychologia*, *48*(4).

Kučera, H., Francis, W., Carroll, J., & Twaddell, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.

Maffre, T., Bregeon, C., Garrigou, C., Barry, I., Raynaud, J. P., & Duvignau, K. (2012). The comprehension of verb-based metaphors by children with pervasive developmental disorder (PDD): A marker of lexical rigidity? *Neuropsychiatrie de l'Enfance et de l'Adolescence*, *60*(5).

Mei, J.-J., Zheng, Y.-M., Gao, Y.-Q., & Yin, H.-X. (1984). *Tongyici cilin*. Shangai: Commercial Press.

Méligne, D., Fossard, M., S., B., Moreaud, O., Duvignau, K., & Démonet, J. (2011). Verb production during action naming in semantic dementia. *Journal of Communication Disorders*, *44*, 379-391.

Morais, A., Olsson, H., & Schooler, L. (2013). Mapping the structure of semantic memory. *Cognitive Science*, *27*(1), 125-145.

Morrison, C., Chappell, T., & Ellis, A. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology Section A*, *50*(3), 528-559.

Nelson, D., McEvoy, C., & Schreiber, T. (2004). The University of South Florida free association rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402-407.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, *45*, 167-256.

Ploux, S., & Victorri, B. (1998). Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes. *Traitement Automatique des Langues*, *39*(1), 161-182.

Schaefer, R. P. (1979). Child and adult verb categories. *Kansas Working Papers in Linguistics*, *4*, 61-76.

Stewart, G. (1994). *Perron-Frobenius theory: a new proof of the basics* (Tech. Rep.). College Park, MD, USA.

Steyvers, M., & Tenenbaum, J. B. (2005). The Large Scale Structure of Semantic Networks: Statistical Analyses and Model of Semantic Growth. *Cognitive Science*, *29*, 41-78.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speaker's early vocabularies. *Developmental Psychology*, *2*(3), 492-504.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440-442.