

Simulating Language Dynamics by Means of Concept Reasoning

Gonzalo A. Aranda-Corral¹, Joaquín Borrego-Díaz², and Juan Galán-Páez²

¹ Department of Information Technology, Universidad de Huelva,
Crta. Palos de La Frontera S/n., 21819 Palos de La Frontera, Spain

² Department of Computer Science and Artificial Intelligence, Universidad de Sevilla,
Avda. Reina Mercedes S/n., 41012 Sevilla, Spain
juangalan@us.es

Abstract. A problem in the phenomenological reconstruction of Complex Systems (CS) is the extraction of the knowledge that elements playing in CS use during its evolution. This problem is important because such a knowledge would allow the researcher to understand the global behavior of the system [1, 2]. In this paper an approach to partially solve this problem by means of Formal Concept Analysis (FCA) is described in a particular case, namely Language Dynamics. The main idea lies in the fact that global knowledge in CS is naturally built by local interactions among agents, and FCA could be useful to represent their own knowledge. In this way it is possible to represent the effect of interactions on individual knowledge as well as the dynamics of global knowledge. Experiments in order to show this approach are given using WordNet.

1 Introduction

Complex System (CS) is a broad concept which has specific features but covers very different systems, with an astonishing variety of dynamics. Among them, particularly interesting are those related with human (rational) activities, as for example, organizations, communities and cities. It is usual to study and simulate these kinds of systems by reducing human behavior to simple (but essential) processes. A traditional methodology in CS research is to model these by designing local interactions between nodes (agents) of the CS, then -by means of simulations- global properties are studied (by checking their reliability, accuracy and validity). Several types of agents interactions can be considered: games, communication (messages), competition, etc.

Particularly interesting is the study of Language Dynamics (LD), a rapidly growing field in CS community, that focuses on all processes related with emergence, evolution, change and extinction of languages [11]. For instance, an approximation to the study of self-organization and evolution of the language

and its semantics to consider the community of users as a CS that collectively build the semantics features of their lexicon. Formal Concept Analysis (FCA) [6] aims to collect formal concepts defined by attributes. Thus FCA tools can be applied to enhance models in order to study LD, in which implicit semantic structures in agent’s language can be considered.

1.1 Naming Games

A popular approach in LD is to model agents’ interaction by means of naming games [19,20]. Naming games were created to explore self-organization in LD (emergence of vocabularies, that is to say, the mapping between words and meanings). Naming games consist on the interaction between two agents, a speaker and a listener. From the basic model, a number of variants for several and specific models can be considered. The aim of the agent community is to achieve a common vocabulary. The minimal naming game is as follows. Each agent has its own context (object/word) and interacts according to the following steps [11]:

1. The speaker selects an object from the current context.
2. The speaker retrieves a word from its inventory associated with the chosen object, or, if its inventory is empty, invents a new word.
3. The speaker transmits the selected word to the listener.
4. If the listener has the word named by the speaker in its inventory and that word is associated with the object chosen by the speaker, the interaction is a success and both players maintain in their inventories only the winning word, deleting all the other words that fitted the same object.
5. If the listener does not have the word named by the speaker in its inventory, or the word is associated to a different object, the interaction is a failure and the listener updates its inventory by adding an association between the new word and the object.

Naming games have been considered both in non-situated and situated models. Situated models place agents in an artificial world, where environmental features as distance between agents or agent’s neighborhood can be considered. Situated models are very interesting because the communication among agents does not obey purely random selections: communication takes place between agents which are able to do it (for example, between neighbor ones). Moreover, naming games with spatially distributed agents allow to model the emergence of different language communities by stabilization of the system [18]. This is due to the fact that the “success” of a linguistic innovation is dependent on whether the group, as a whole, has adopted it or not.

Our interest in naming games is based on their *adaptive nature*, that is to say, naming games can produce changes in the lexicon of both, the speaker and listener, as side effect. Thus, agent’s lexicon changes during its live within the system. By considering the possible results in each step of the naming game we can list the side effects on agents’ lexicon on the game (see [18]).

From FCA point of view, minimal game is a very inspiring interactive method in which it can be applied. For example, by considering *synsets* (sets of cognitive

synonyms) from *WordNet*¹ as attributes, which provides an implicit meaning. There exist several variants of the above interaction by considering different levels of reasoning ability of agents.

1.2 Reasons for Selecting FCA (and WordNet)

Classic LD does not consider strong semantic features on agents' interactions. FCA provides a general framework in which semantic features can be added to LD (at object/attribute level). It is interesting to consider a real case study, a language with coherent semantics (also compatible with FCA) as emergence target, to discuss this proposal. In this paper WordNet is selected as the case study but of course it can also be used with any other object-attribute system.

Aim of the Paper. The aim is to describe how the concept lattice associated to the full language emerges (in asymptotic terms) from a community of agents by means of FCA-based semantic interactions. Specifically, the aim is to describe a number of experiments with FCA-based variants of LD approach using *naming games-based models*, by showing models' behaviors. In this way we demonstrate that FCA can enrich LD models, particularly those that focus on vocabulary emergence.

To illustrate the different proposals, WordNet (subsets of) is considered as an universal vocabulary (a global knowledge which the multiagent system aim to achieve). In this way experiments can be confronted with the real structure of a vocabulary (although models presented in this paper -a first approach- are basic and they do not consider every WordNet feature).

The selection of an existent lexical database is not arbitrary. Human languages are surprisingly robust and stable, with persistent categories [14]. In WordNet terms, synsets are categorical classifications that have emerged from human interaction, thus it is a good candidate in order to model its emergence.

Structure of the Paper. In the next section some remarks on FCA and its use in LD modeling are considered. Section 3 presents some variants of naming games by using FCA elements. In order to perform experiments, WordNet lexicon database is selected. Section 4 is devoted to describe a FCA-based version of the naming game. In Sect. 5 we analyse an interesting variant of naming games which uses concept reasoning (actually reasoning with implications of agents' contexts and attribute exploration idea) to perform the interaction. This variant shows a behavior similar than the one in Steels' spatially distributed naming games [18], thus in order to simulate interlingua phenomena, hybrid games are presented in Sect. 6. Section 7 is devoted to discuss experimental results and, in Sect. 8, some conclusions of the work done are given.

¹ <http://wordnet.princeton.edu>

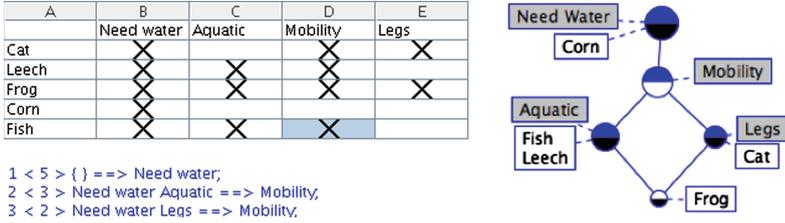


Fig. 1. Formal context, associated concept lattice and Stem Basis

2 Background: Formal Concept Analysis and Implications

According R. Wille, FCA mathematizes the philosophical understanding of a concept as a unit of thoughts composed of two parts: the extent and the intent [6]. The extent covers all objects belonging to this concept, while the intent comprises all common attributes valid for all the objects under consideration. It also allows the computation of concept hierarchies from data tables. In this section, we succinctly present basic FCA elements, although it is assumed that the reader is familiar with this theory (the fundamental reference is [6]).

A formal context is represented as $M = (O, A, I)$, which consists of two sets, O (objects) and A (attributes) and a relation $I \subseteq O \times A$. Finite contexts can be represented by a 1-0-table (representing I as a Boolean function on $O \times A$). See Fig. 1 for an example of formal context about living beings.

The FCA main goal is the computation of the concept lattice associated with the context. Given $X \subseteq O$ and $Y \subseteq A$ it defines

$$X' := \{a \in A \mid oIa \text{ for all } o \in X\} \text{ and } Y' := \{o \in O \mid oIa \text{ for all } a \in Y\}$$

A (formal) concept is a pair (X, Y) such that $X' = Y$ and $Y' = X$. For example, concepts from formal context about living beings (Fig. 1, left) are depicted in Fig. 1, right. Actually in Fig. 1, each node is a concept, and its intension (or extension) can be formed by the set of attributes (or objects) included along the path to the top (or bottom). E.g. The node tagged with the attribute Legs represents the concept $(\{Legs, Mobility, NeedWater\}, \{Cat, Frog\})$.

2.1 Implications and Basis

Logical expressions in FCA are *implications between attributes*, pair of sets of attributes, written as $Y_1 \rightarrow Y_2$, which is true with respect to $M = (O, A, I)$ according to the following definition. A subset $T \subseteq A$ *respects* $Y_1 \rightarrow Y_2$ if $Y_1 \not\subseteq T$ or $Y_2 \subseteq T$. It says that $Y_1 \rightarrow Y_2$ holds in M ($M \models Y_1 \rightarrow Y_2$) if for all $o \in O$, the set $\{o\}'$ respects $Y_1 \rightarrow Y_2$. See [5, 6, 17] for more information.

Definition 1. Let \mathcal{L} be a set of implications and L an implication of M .

1. L follows from \mathcal{L} ($\mathcal{L} \models L$) if each subset of A respecting \mathcal{L} also respects L .
2. \mathcal{L} is complete if every implication of the context follows from \mathcal{L} .
3. \mathcal{L} is non-redundant if for each $L \in \mathcal{L}$, $\mathcal{L} \setminus \{L\} \not\models L$.
4. \mathcal{L} is a basis for M if it is complete and non-redundant.

It can obtain a basis from the *pseudo-intents* [8] called *Stem Basis* (SB). SB is only an example of a implication basis. In this paper none specific property of the SB can be used, so it can be replaced by any other basis.

In order to work with formal contexts, stem basis and association rules, the Conexp² (cf. [21]) software has been selected. It has been used as a library to build the component which provides implications (and association rules) to the reasoning module of the system we have used in several applications of FCA.

2.2 A Formal Context Associated to WordNet

As it was mentioned before, it is interesting to consider a real and structured language in order to exploit FCA semantic features, thus in this paper experiments are performed on subsets of WordNet system. A Formal context associated to WordNet has been considered by using words as objects of the context and synsets as attributes. Synsets are sets of synonymous words, thus they can be considered as a potential definition (meaning) of each word. The concept lattice associated to this huge lexical database can not be computed, thus small subsets of WordNet have been taken instead, in order to be able to compute FCA elements (concept lattice and SB) in short time. In this way, it will be possible to evaluate the soundness of the proposed models with respect to the FCA elements associated to these subsets.

2.3 Formal Contexts Associated to Agents in Complex Systems

Actually two different scopes of formal contexts should be considered in the experiments. The first one is the formal context associated to the whole language considered, the **global knowledge**, and the second one is the formal context associated to the lexicon that each agent owns, the **individual knowledge**. Starting from an initial lexicon (usually randomly selected) for each agent, they interact using variants of naming games and the result of these interaction transforms their contexts (so their formal concepts).

Finally, to compute the collective knowledge, the individual knowledge of each agent is aggregated to obtain a *similarity matrix*. The entries of this matrix is the number of agents owning each pair object-attribute. A pair belongs to the **collective knowledge** only if its value in the similarity matrix is above a certain collective knowledge threshold CK_{th} (which will be detailed later).

Since the aim of the paper is to describe how the concept lattice associated to the full language emerges (in asymptotic terms) from the interaction of a community of agents, it is interesting to study what kind of FCA-based naming

² <http://sourceforge.net/projects/conexp/>

games are efficient to achieve this goal, and how these games output stable communities which need other type of interactions (based on attribute exploration idea, for example) to intercommunicate their lexicons.

2.4 Assumptions on the Model

Our approach uses several assumptions -adapted from those enumerated in [9] to a FCA framework- for each model presented in this paper³:

- *All agents have same semantic space and pre-existing semantic categories.* This assumption is selected when agents “understand” a pair (*word, synset*) equally to other agents that knew the same pair.
- *Agents are equipped with a symbolic communication ability.* In our case, the ability to reason with Stem basis.
- *Agents can read each other’s communicative intentions.* In our case, we simplify this to two performatives (from speech acts): *request* and *answer*.
- *Agents have imitation ability: agents accept information from other agents (credulous agents).*
- *Agents continually detect recurrent patterns.* In our case, agents detect true implications. Our model is not realistic in this assumption, because agents in Sect. 5 selects implications from Stem basis instead of any true implication.
- *Agents have sequencing ability. In our case, sequencing is limited to one-step memory.* We will see that It is sufficient in our model.
- *Agents’ behavior is governed by rule competition.* This idea is implemented by selecting, as the knowledge basis of agents, the stem basis (Sect. 5).

The assumptions seem stronger in some cases, compared with other model assumptions as in [10], where minimal abilities for initial agents are supposed.

2.5 Parameters on the Models

Mainly five parameters are considered (see also [7]):

- N is the population size. The values chosen in the experiments are conditioned by the feasibility of the computation of each model.
- δ is the probability for an agent to have within its initial knowledge a pair *lema-synset* (object-attribute) (see section bellow). As general rule, δ is selected in a value range which provides each pair *lema-synset* to appear in at least one agent from the overall population with probability $P = 0.95$.
- Convergence (stabilization) criteria: The convergence test checks whether every existent pair *lema-synset* is present within the *collective knowledge*.
- The collective knowledge threshold CK_{th} is selected within the range [90%–100%]. This selection has the aim of assuring the convergence to the full language in (almost) every agent.
- The size of the selected subset of WordNet in the experiments is determined by both, the fact that the subsystem must contain complete synsets and by the computational feasibility. The complex concept lattices associated to subsystems of Wordnet are hard to compute [1].

³ Some of them will be weakened in future works.

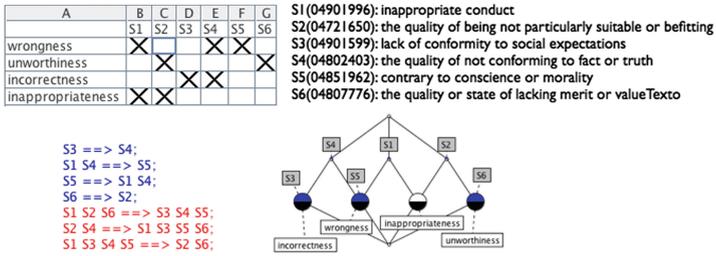


Fig. 2. Example (1st row: context and synsets definitions; 2nd row: stem basis and concept lattice)

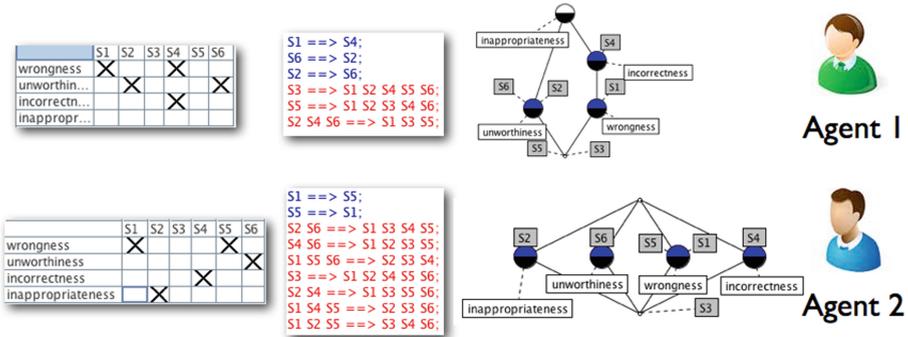


Fig. 3. Agents for the example

3 Modeling Communication as FCA-based Naming Games

Roughly speaking, the goal of using FCA in naming games is to analyze interactions between agents with a partial knowledge in order to study how a collective knowledge emerges. In these communication games, each interaction is a communicative act where two agents interchange new knowledge. In the experiments performed, the creation steps have been dropped because the aim of these models is to induce the emergence of WordNet.

In order to model communication between agents, the English lexical database *WordNet* has been chosen as the global knowledge. In *WordNet* nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms called *Synsets*, where each of these express a distinct concept. *Synsets* are interlinked by means of conceptual-semantic and lexical relations.

In order to illustrate the process, a tiny WordNet subset has been chosen (see Fig. 2) and the initial knowledge of two agents which will perform the communicative act is shown in Fig. 3.

3.1 Communicative Process Outline

The simulation environment for communication games has been considered as a grid (the world) where agents move freely. In each step, if an agent meets another, a communicative act takes place. Roughly speaking it is as follows.

1. The world is randomly initialized for a given density (*population/gridsize*). Each agent starts with an initial knowledge randomly taken from the global one. To obtain successful communicative games, it is necessary that the union of the initial local knowledge of each agent contains approximately all concepts within the global knowledge.

As it was already commented, the probability P for each pair within the selected WordNet subset to appear in the initial local knowledge of at least one agent is given by:

$$P = 1 - (1 - \delta)^N$$

It is suggested to carry out communication games with at least $P > 0.95$, thus the value δ to be considered depends on the number of agents N in the world (i.e. for $N = 200$ it is suggested that $\delta \geq 0.015$).

2. In each time step, each agent moves randomly to an adjacent cell.
3. Each agent (speaker) chooses randomly a listener agent within the agents in the same cell, in order to start a communicative process (request).
4. After the simulation, the collective knowledge can be measured.

There are different ways of performing the communicative process as well as different ways of measuring the collective knowledge. Those will be depicted in the following sections. Due to the huge size of the *WordNet* database, in order to compute the simulations of communication games, different subsets of *WordNet* have been considered as global knowledge. In order to enrich language dynamics within communication games, only connected subsets of *WordNet* and only formed by full *synsets* have been considered.

3.2 Formal Contexts as Agents' Knowledge in Communication Games

In order to work with FCA-based naming games, the **individual knowledge of an agent** is considered as a local Formal Context, in which *lemas* are *objects* and *synsets* are *attributes*. A relation between an object o_i and an attribute a_i means that that the lema o_i belongs to the synset a_i .

4 Modeling Emergence of WordNet by Intent-Extent Games

The first model is the most direct one due to the relative similarity between a *formal concept* and a *synset*. In this communication game, the communicative act consists on direct interchange of lemmas (objects) and synsets (attributes) between the speaker and the listener.

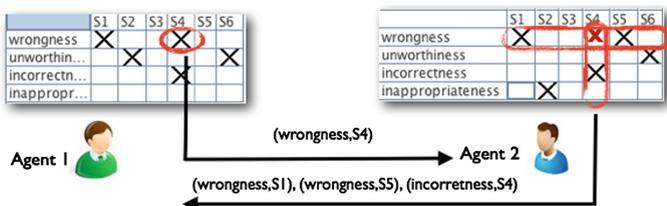


Fig. 4. Intent-extent communicative act

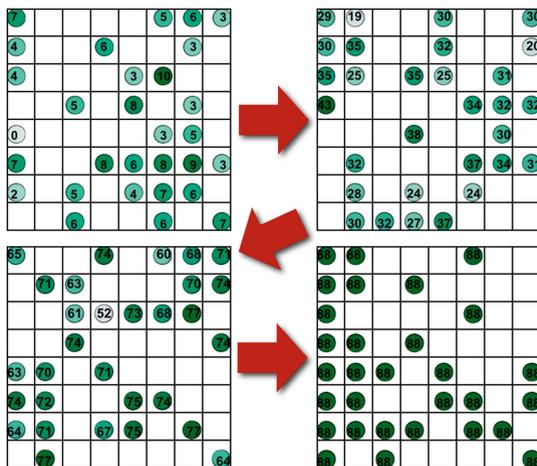


Fig. 5. Evolution of agents knowledge within the world grid

Communicative Act. Each time step, the speaker randomly chooses a pair (o_i, a_j) from its local knowledge (formal context) and sends it (request) to the listener. The answer of the listener will be the two sets $intent(o_i)$ and $extent(a_j)$ (relative to its own formal context). In case the listener does not have any information about o_i or a_i , it will returns an empty set and will add the pair (o_i, a_j) to its local knowledge (see Fig. 4).

Collective Knowledge Emergence. In order to detect and measure the emerging knowledge due to agents interactions in this communication game, the *collective knowledge* is obtained by computing the *similarity matrix*. In each time step the *error rate* between the global and the collective knowledge is measured as the difference between the collective and the global knowledge.

In Fig. 5 four different states of the communicative process for this model are shown. A small example has been chosen in order to show a representation of the world. The circles are agents which randomly move within the grid. The number on the agents shows the number of complete synsets they have within their *individual knowledge*.

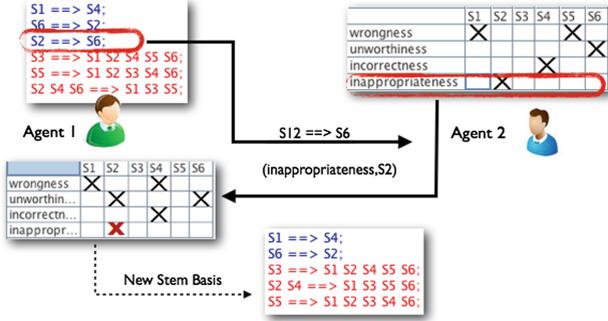


Fig. 6. Communicative act in stem basis game

Convergence Criteria. The communication game ends when the *collective knowledge* emerged from agents interactions is equal to the *global knowledge*. It is worthy to note that the convergence rate of the game highly depends on the collective knowledge threshold CK_{th} considered (see Sect. 7 below).

5 Modeling Emergence of WordNet by Stem Basis Games

This model aims to exploit the power of stem basis (SB) in knowledge detection tasks. In a first approach the communication process goal is the emergence of the collective knowledge by detecting and eliminating inconsistencies within local knowledge of agents. The communicative process in this case concerns to consistency questions. Each agent will contrast its knowledge with others', in order to detect and fix inconsistencies (see Fig. 6).

Communicative Act. The speaker computes the SB of its local formal context, randomly chooses a rule r_i from it and sends it (request) to the listener. If r_i is true within the listeners' local knowledge, it returns a positive answer and finish the communicative act. Otherwise it returns a negative answer and sends to the speaker a counter example (o_i) for r_i , and the speaker adds it to its local knowledge in order to fix the inconsistency. This communicative act is similar to one step of the *attribute exploration* (cf. [6]).

Collective Knowledge Emergence. In this case, as the model works with SB, the collective knowledge has to be considered as the *collective consistent knowledge*, that is to say, the SB corresponding to the collective knowledge have to be consistent with the SB corresponding to the global knowledge. As the rules of the game have changed, in this case, the collective knowledge is measured by its consistency. In order to estimate the soundness of the emerged knowledge, it should be verified whether the true implications within the collective knowledge

are entailed by the SB of the original vocabulary. Firstly, the *similarity matrix* is computed as in the former model, and a collective formal context is obtained by filtering de similarity matrix with the aforementioned collective knowledge threshold CK_{th} . Then the collective SB is computed and its consistency is verified against the global SB.

Convergence Criteria. This communication game ends when the game reach the equilibrium, that is to say, when there are no more inconsistencies between agents local knowledge. It is worthy to note that from this model does not emerges the global knowledge (it is not the aim), but a knowledge consistent with the global one. In the following section an hybrid model will be considered in order to get both, consistency and completeness.

6 Modeling Emergence of WordNet by Hybrid Games

In order to obtain a better model for language emergence, an hybrid model is considered. Particularly, to complete the stem basis game, which stabilizes before agents' local knowledge converges to the global knowledge.

The consistency based approach is interesting but does not provide full emergence of collective knowledge. Thus in this hybrid approach the two communicative act types depicted above will be considered, one based on consistency (stem basis interactions) and the second based on direct information exchange (intent-extent interactions). In this model, another question arises, how to merge both types of communicative act?

Communicative Act. In a first approach of merging both types of communicative acts, the simplest solution is to use one communication type or another with a certain probability P (usually $P = 0.5$). This first approach produces a behavior very similar to the intent-extent game, thus it is not very interesting, and another approach should be considered.

Firstly, two new parameters should be considered. One is the time period τ_{eq} necessary to consider that the game as arrived to a equilibrium state, thus, every inconsistency between agents' local knowledge have been suppressed. That is to say, the time period in which agents' local knowledge does not changes. The second parameter is the time period $\tau_{int-ext}$ in which agents interchange information freely (as in the first model) in order to enrich their local knowledge, but leading to new inconsistencies. The process for this second approach is as follows.

Agents normally communicate others as in they do in the stem basis game. If after a time period τ_{eq} the system stays in equilibrium, agents' communication type changes to the one of the intent-extent game during a time period $\tau_{int-ext}$. Then they come back to behave as in the stem basis game in order to solve inconsistencies until the next equilibrium state.

Collective Knowledge Emergence. In this case the both notions of collective knowledge above mentioned should be considered, in order to evaluate both, consistency an knowledge emergence.

Convergence Criteria. The game ends when both objectives are reached, the emergence of consistency and the global knowledge, within the collective agents knowledge.

7 Experiments and Discussion

In order to study the convergence of agents' collective knowledge for each of the aforementioned communication games, many experiments with the different models have been carried out. A connected subset of *WordNet* (as it was mentioned before) has been selected to be used in all experiments.

In Fig. 8 the results of some of those experiments are shown. The figures show the emergence of collective knowledge which tends to the global knowledge (presented as a percentage of the global knowledge). In the cases which correspond, it is also shown the evolution of the implications, associated to the collective knowledge, with respect to those associated to the global one.

The subset of *WordNet* considered for the experiments is of relatively small size (around 400 lema-synset pairs) due to the high computation time of the

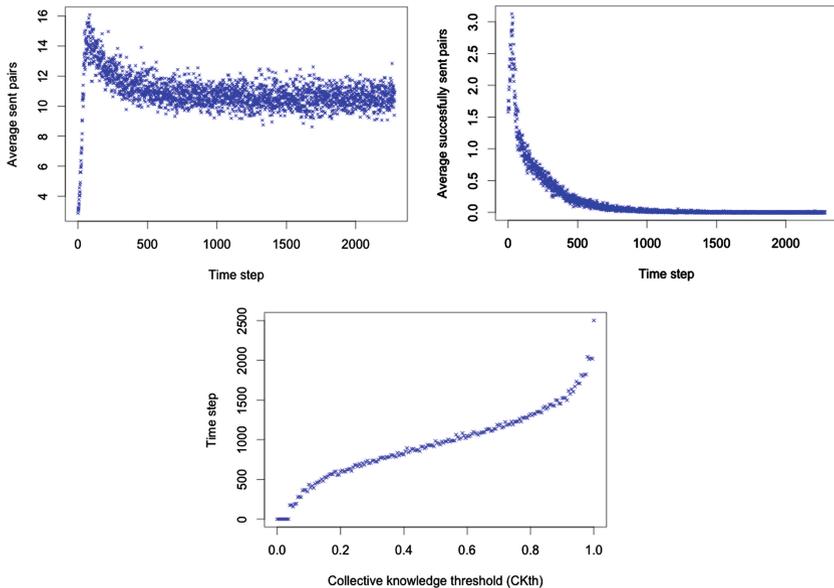


Fig. 7. First row: average number of pairs (lemma-synset) involved in communicative act (left) and the average number of positive (new pair for the listener) ones (right). Second row: convergence rate with respect to the collective knowledge threshold CK_{th}

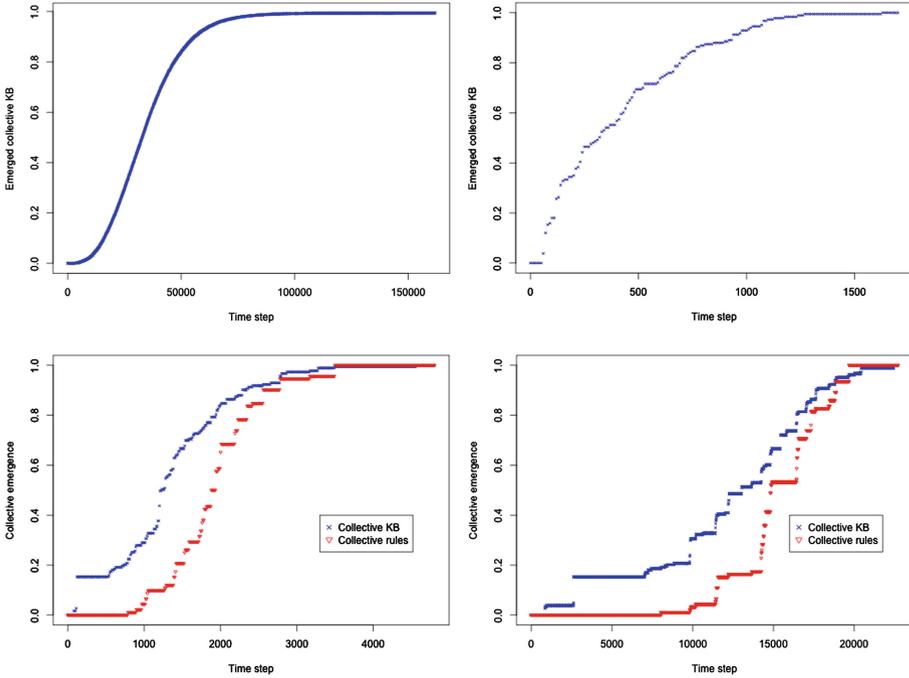


Fig. 8. Convergence of collective Knowledge in communication games based on intent-extent (first row), and hybrid games (second row).

Stem Basis for huge formal contexts. Thus, in order to show the fact that agents' collective knowledge converges asymptotically towards the global knowledge, an experiment for the intent-extent game has been performed using a huge subset of *WordNet* (around 26700 lema-synset pairs) (see Fig. 8 top left). The rest of the figures corresponds to intent-extent game using the same *WordNet* subset than in the others (Fig. 8 top right), hybrid probability-based game with $P = 0.5$ (Fig. 8 bottom left) and hybrid rounds-based game (Fig. 8 bottom right).

It should be noted that no experimental results are shown for the game based exclusively in Stem Basis, due to the fact that the system stabilizes before any concept exceed the collective knowledge threshold (CK_{th}) (which should be high) in order to be considered as collective knowledge. This phenomena is similar to others in LD simulation (language competing).

In the plot corresponding to the hybrid rounds-based game (Fig. 8 bottom right), it is very interesting to observe the different behaviors that arises depending on the communication type being used. When the communication game is in a period of intent-extent communications, intervals of stable knowledge appears (see also Fig. 8, bottom right). Recall that the intent-extent game consists on random knowledge interchanges and introduces *inconsistencies* (counterexamples). In the other hand, when the communication game is in a period of Stem

Basis communications, the knowledge base slightly increases (due to the counterexamples interchange) but the number of collective implications consistent with the global knowledge increases significantly. That is to say, it introduces *consistency* to the knowledge acquired in the previous period.

In order to show the evolution of communicative acts, Fig. 7 depicts the average number of pairs (lemma-synset) that are sent in each time step (1st row, left), and the number of successful ones (1st, right), both for the intent-extent experiment. Finally, Fig. 7 (2nd row) shows the experimental relationship between the collective knowledge threshold (CK_{th}) and average convergence time step. In a nutshell, intend-extend based communication produces collective knowledge quite fast, but it is inconsistent (with respect to the global knowledge) until the whole global knowledge has been learnt. While the communication based on SB produces collective knowledge and don't converges by itself to the global knowledge, but the partial knowledge produced is consistent with the global one.

8 Conclusions and Future Work

In this paper it is shown how to apply FCA in order to enrich LD simulation, in the particular case of naming games in Language Dynamics. FCA provides a solid formal semantic characterization of implicit conceptual structures of Language users. In this way it is possible to analyze the semantic evolution of LD. For instance, a promising research task is to model qualitative *category games* [11].

A consequence of the results of this work is that language convergence seems to be governed by the shared vocabulary (the mapping between words and meaning) instead of the shared language. This conclusion can be justified because the selected δ (which estimates the size of initial vocabulary of agents) is small, and with it the number of pairs (*word, synset*) initially in each agent. It remains to be investigated vocabulary distributions distinct than the uniform one, for example those that take into account the agents within agent's neighborhood (in order to empower the sharing of vocabulary among close agents).

An interesting research line is to consider weighted distributions on the vocabulary (and concepts [4]), in order to compare local and global knowledge. Also it would be interesting to consider non-uniform distributions of agents within the world (cells). Other distributions (in big sized worlds) would induce the existence of sub-communities with different languages (speaker communities). In the future will also be considered the case where languages get in touch due to communication between speaker communities. This second phenomena is only possible to be simulated if the pair speaker-listener behavior is not driven by means of purely random walks and distributions. In this case, it occurs that this mediates the communication by topological features [18] (see also [13]). The analysis of agent lexicons -in preliminary experiments- reveals that agents will develop stable knowledge within their cluster, but will also develop a second language, an interlingua which is weaker but shared among different clusters [18].

With respect to the use of significative vocabulary subsets (from WordNet or Thesaurus), it is interesting the use of *data weeding* techniques [15] to classify

which of them can be explained by CS emergence. The discovering of LD models to explain this would show new ideas on how such vocabularies emerge in the real world. In the case of the concept lattice provided, it is interesting to study the emergence of *concept neighborhoods* [16]. Also, we are investigating other semantic relationships distinct than synsets.

Lastly, it can be sound to use association rules (subsets of Luxemburger basis [12]) instead of Stem Basis. This choice is very related with the idea of lexicon mediated by confidence in the relationship. The reasoning with association rules is more complex, although promising approaches can be exploited [3].

References

1. Aranda-Corral, G.A., Borrego-Díaz, J., Galán-Páez, J.: Scale-free structure in concept lattices associated to complex systems. In: Proceedings of International Conference on Complex Systems, ICCS 2012 (2012) (to appear in IEEE press)
2. Aranda-Corral, G.A., Borrego-Díaz, J., Galán-Páez, J.: Complex concept lattices for simulating human prediction in sport. *J. Syst. Sci. Complex.* **26**(1), 117–136 (2013). (to appear)
3. Balcázar, J.L.: Redundancy, deduction schemes, and minimum-size bases for association rules. *Log. Meth. Comput. Sci.* **6**(2), 1–23 (2010)
4. Belohlavek, R., Macko, J.: Selecting important concepts using weights. In: Jäschke, R. (ed.) ICFCA 2011. LNCS (LNAI), vol. 6628, pp. 65–80. Springer, Heidelberg (2011)
5. Bertet, K., Monjardet, B.: The multiple facets of the canonical direct unit implicational basis. *Theor. Comput. Sci.* **411**(22–24), 2155–2166 (2010)
6. Ganter, B., Wille, R.: *Formal Concept Analysis - Mathematical Foundations*. Springer, Berlin (1999)
7. Gong, T., Wang, W.S.-Y.: Computational modeling on language emergence: a coevolution model of lexicon, syntax and social structure. *Lang. Linguist.* **6**, 1–42 (2005)
8. Guigues, J.-L., Duquenne, V.: Familles minimales d’implications informatives resultant d’un tableau de données binaires. *Math. Sci. Hum.* **95**, 5–18 (1986)
9. Ke, J., Holland, J.H.: Language origin from an emergentist perspective. *Appl. Linguist.* **27**(4), 691–716 (2006)
10. Kirby, S.: Syntax without Natural Selection: How compositionality emerges from vocabulary in a population of learners. In: Knight, C. (ed.) *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pp. 303–323. Cambridge University Press, Cambridge (2000)
11. Loreto, V., Baronchelli, A., Mukherjee, A., Puglisi, A., Tria, F.: Statistical physics of language dynamics. *J. Stat. Mech. Theory Exp.* **4**, 1–29 (2011)
12. Luxemburger, M.: Implications partielles dans un contexte. *Math. Inf. Sci. Hum.* **113**, 35–55 (1991)
13. Patriarca, M., Castelló, X., Uriarte, J.R., Eguiluz, V.M., San, M.: Miguel, modelling two-language competition dynamics. *Adv. Complex Syst.* **15**(3–4), 1–24 (2012)
14. Mukherjee, A., Tria, F., Baronchelli, A., Puglisi, A., Loreto, V.: Aging in language dynamics. *PLoS ONE* **6**(2), e16677 (2011). doi:[10.1371/journal.pone.0016677](https://doi.org/10.1371/journal.pone.0016677)

15. Priss, U., Old, L.J.: Data weeding techniques applied to roget's thesaurus. In: Wolff, K.E., Palchunov, D.E., Zagoruiko, N.G., Andelfinger, U. (eds.) KONT 2007 and KPP 2007. LNCS (LNAI), vol. 6581, pp. 150–163. Springer, Heidelberg (2011)
16. Priss, U., Old, L.J.: Concept neighbourhoods in lexical databases. In: Kwuida, L., Sertkaya, B. (eds.) ICFCA 2010. LNCS (LNAI), vol. 5986, pp. 283–295. Springer, Heidelberg (2010)
17. Rudolph, S., Völker, J., Hitzler, P.: Supporting lexical ontology learning by relational exploration. In: Priss, U., Polovina, S., Hill, R. (eds.) ICCS 2007. LNCS (LNAI), vol. 4604, pp. 488–491. Springer, Heidelberg (2007)
18. Steels, L., McIntyre, A.: Spatially distributed naming games. *Adv. Complex Syst.* **1**(4), 301–323 (1999)
19. Steels, L.: A self-organizing spatial vocabulary. *Artif. Life* **2**(3), 319–332 (1995)
20. Steels, L.: Self-organizing vocabularies. In: *Artificial Life V: Proceedings of 5th International Workshop on the Synthesis and Simulation of Living Systems*, pp. 179–84. MIT Press (1996)
21. Yevtushenko, S.A.: System of data analysis “Concept Explorer”. In: *Proceedings of 7th National Conference on Artificial Intelligence KII-2000*, pp. 127–134 (2000)