



## Research Networking Programmes

Short Visit Grant  or Exchange Visit Grant

*(please tick the relevant box)*

### Scientific Report

**Scientific report (one single document in WORD or PDF file) should be submitted online within one month of the event. It should not exceed eight A4 pages.**

**Proposal Title:** Lexical acquisition in bilingual contexts: aspects of computational modelling

**Application Reference N°:** 5604

#### 1) Purpose of the visit

The visit was framed within grant proposal 5604, whose main objective is to explicitly address computational issues in modelling lexicon organisation, by focusing on linguistic information from the input signal and sentence context. The proposal is tightly coupled with another companion proposal, presented by Claudia Marzi (ILC-CNR), “Lexical acquisition in bilingual contexts: aspects of (extra)linguistic and psycholinguistic modelling”, which is intended to focus on the cognitive and extra-linguistic factors involved in bilingual word recognition, with a view to exploring implications, requirements and constraints on computational models of bilingual acquisition.

The main goal of the present visit was to investigate algorithmic and computational properties of lexical organisation which are mainly based on the linguistic nature of training data: phonotactic organisation, underlying morphological structure, type and frequency distribution of lexical families such as inflectional paradigms, derivational/semantic word clusters and form neighbourhoods. The general goal is articulated into the following subsidiary goals: i) investigate modelling requirements of aspects of bilingual lexical organisation, ii) review compliance with these requirements of a selection of symbolic/subsymbolic algorithms/tools for morphology learning, iii) envisage possible improvements which can make current machine learning algorithms more amenable to meeting these requirements, iv) assess their impact in simulating word processing, access and acquisition.

## **2) Description of the work carried out during the visit**

The two full day meeting was held at the Computational Linguistics and Psycholinguistics Research Centre of the Linguistics department of the faculty of Arts of the University of Antwerp, headed by Prof. Walter Daelemans. Walter Daelemans, Emmanuel Keuleers (Ghent University) and Vito Pirrelli (ILC-CNR) attended the meeting for its entire duration.

The first part of my visit was devoted to presentation of some preliminary results of differential dynamics in L1/L2 acquisition, computationally simulated by running incremental Hebbian Self-Organising Maps (Ferro et al., 2011; Marzi et al., 2012) under different training conditions in terms of language-pairs, age and time of L1/L2 exposure. Simulations showed non trivial evolutionary patterns in terms of accuracy in activation and recall on both acquired and novel words in L1/L2, number of processing nodes recruited for the two languages, average error rate of acquired inter-node connections, plasticity of the maps in developing novel connections or parasitically exploiting existing ones.

After extensive discussion of the significance of these results in the light of the recent literature on language-independent access to the bilingual lexicon (Dijkstra & Van Heuven, 2001), we considered opportunities for further testing and cross-checking of their consistency/stability. Furthermore, the discussion touched upon computational and experimental issues. First, we discussed the possibility of using cluster computing as a way to scale up the limitations of current computer architectures and run simulations on more realistic lexical data sets. We then considered using Hebbian Self-Organising maps as models of peripheral, modality-dependent, lexical access, intended to recode input words through levels of distributed activation patterns over processing nodes (activation chains). Activation chains can then be provided as input representations to a memory-based system (Daelemans & van den Bosch, 2007) addressing a morphological classification task (e.g. assigning the appropriate inflection class to singular German noun forms). Finally, we discussed the possibility of using measures of activation over map chains to test hypotheses about perceived levels of wordlikeness by native speakers on a shared data set (Bailey & Hahn, 2001; Hahn & Bailey, 2005).

## **3) Description of the main results obtained**

A dataset of German nouns was selected and shared for training Hebbian Self-Organising Maps and testing them on a pluralisation task. A protocol for exchanging activation chains recoding input words was agreed upon. Due to current hardware limitations, it was also agreed to run a preliminary set of experiments on a subset of 500 German noun forms only. An experimental protocol was designed for the purpose.

Concerning the perceived wordlikeness task, we designed a distributed representation protocol for phonological segments in terms of acoustic and articulatory features, and recoded Baileys and Hahn's experimental data set with the designed protocol. Several measures approximating the notion of wordlikeness were entertained (e.g. mean of integrated activation pattern, mean of BMUs activation, gmean/mean of connections between BMUs). They will be subjected to extensive testing.

## **4) Future collaboration with host institution (if applicable)**

It was commonly felt that all three groups can considerably benefit from joint work on shared objectives. This is likely to be conducive to an overall assessment of the impact and implications of different machine-learning algorithms and biologically inspired simulative frameworks on understanding human behavioural evidence for word processing, access and acquisition. This work is bound to have an impact on the design of more complex architectures modelling both L1 and L2 lexical processing from a unifying perspective.

**5) Projected publications / articles resulting or to result from the grant (*ESF must be acknowledged in publications resulting from the grantee's work in relation with the grant*)**

It was agreed to report results of planned experiments at the forthcoming NetWordS Workshop (Dubrovnik 19-20 September 2013). We also considered publication of more mature results in major international journals. More careful planning will be possible after analysis of preliminary results.

**6) Other comments (if any)**

A full list of references is enclosed as Attachment A.

## References

- Bailey T.M., Hahn U. (2001) Determinants of Wordlikeness: Phonotactics or Lexical Neighborhoods? *Journal of Memory and Language*, 44, 568-591.
- Daelemans W., van den Bosch A. (2005) *Memory-Based Language Processing*. Cambridge, UK: Cambridge University Press.
- Hahn U., Bailey T.M. (2005) What makes words sound similar? *Cognition*, 97, 227-267.
- Dijkstra T., Van Heuven W.J.B.(2002) The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5 (3), 175-197.
- Ferro M., Marzi C., Pirrelli V. (2011), A Self-Organizing Model of Word Storage and Processing: Implications for Morphology Learning. *Lingue e Linguaggio*, vol. X(2), 209 - 226. Bologna, Il Mulino.
- Marzi C., Ferro M., Pirrelli V. (2012), Word alignment and paradigm induction. In Marzi C., Pirrelli V. (eds.) *Lingue e Linguaggio*, vol. XI (2) 251 - 274. Bologna, Il Mulino.