
Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016

5-6 December 2016, Napoli

Anna Corazza, Simonetta Montemagni and Giovanni Semeraro (dir.)

DOI: 10.4000/books.aaccademia.1666

Publisher: Accademia University Press

Place of publication: Torino

Year of publication: 2016

Published on OpenEdition Books: 26 July 2017

Serie: Collana dell'Associazione Italiana di Linguistica Computazionale

Electronic ISBN: 9788899982546



<http://books.openedition.org>

Printed version

Number of pages: 317

Electronic reference

CORAZZA, Anna (ed.) ; MONTEMAGNI, Simonetta (ed.) ; and SEMERARO, Giovanni (ed.). *Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016: 5-6 December 2016, Napoli*. New edition [online]. Torino: Accademia University Press, 2016 (generated 21 December 2020). Available on the Internet: <<http://books.openedition.org/aaccademia/1666>>. ISBN: 9788899982546. DOI: <https://doi.org/10.4000/books.aaccademia.1666>.

© Accademia University Press, 2016

Creative Commons - Attribution-NonCommercial-NoDerivatives 4.0 International - CC BY-NC-ND 4.0

**Proceedings
of the
Third Italian Conference
on
Computational Linguistics
CLiC-it 2016**

5-6 December 2016, Napoli

Editors:

**Anna Corazza
Simonetta Montemagni
Giovanni Semeraro**



aA

CLiC-it



© 2016 by AILC - Associazione Italiana di Linguistica Computazionale
sede legale: c/o Bernardo Magnini, Via delle Cave 61, 38122 Trento
codice fiscale 96101430229
email: info@ai-lc.it

Pubblicazione resa disponibile
nei termini della licenza Creative Commons
Attribuzione – Non commerciale – Non opere derivate 4.0



Accademia University Press
via Carlo Alberto 55
I-10123 Torino
info@aAccademia.it

isbn 978-88-99982-08-9
www.aAccademia.it/CLIC_2016

Accademia University Press è un marchio registrato di proprietà
di LEXIS Compagnia Editoriale in Torino srl

Contents

Andrea Abel, Aivars Glaznieks, Lionel Nicolas and Egon Stemle An extended version of the KoKo German L1 Learner corpus	13
Linda Alfieri and Fabio Tamburini (Almost) Automatic Conversion of the Venice Italian Treebank into the Merged Italian Dependency Treebank Format.....	19
David Alfter and Yuri Bizzoni Hybrid Language Segmentation for Historical Documents.....	24
Anita Alicante, Anna Corazza, Francesco Isgrò and Stefano Silvestri Relation mining from clinical records.....	29
Anita Alicante, Anna Corazza and Antonio Pironti Twitter Sentiment Polarity Classification using Barrier Features.....	34
Daniela Baiamonte, Tommaso Caselli and Irina Prodanof Annotating Content Zones in news articles	40
Gianni Barlacchi, Azad Abad, Emanuele Rossinelli and Alessandro Moschitti Appetitoso: A Search Engine for Restaurant Retrieval based on Dishes	46
Pierpaolo Basile, Valerio Basile, Elena Cabrio and Serena Villata Argument Mining on Italian News Blogs.....	51
Pierpaolo Basile, Annalina Caputo, Roberta Luisi and Giovanni Semeraro Diachronic Analysis of the Italian Language exploiting Google Ngram	56
Giulia Benotto, Emiliano Giovannetti and Simone Marchi Investigating the Application of Distributional Semantics to Stylometry.....	61
Toine Bogers, Georgeta Bordea, Paul Buitelaar, Nicola Ferro and Gianmaria Silvello IR Scientific Data: How to Semantically Represent and Enrich Them.....	66
Stavros Bompolas, Marcello Ferro, Claudia Marzi, Franco Alberto Cardillo and Vito Pirrelli Reassessing inflectional regularity in Modern Greek conjugation	72
Roberto Bottini, Daniel Casasanto, Andrea Nadalini and Davide Crepaldi Stepping out of the Chinese Room: Word meaning with and without consciousness.....	78
Alice Bracchi, Tommaso Caselli and Irina Prodanof Enriching the Ita-TimeBank with Narrative Containers.....	83
Valeria Caruso, Anna De Meo and Vincenzo Norman Vitale Increasing information accessibility on the Web: a rating system for specialized dictionaries.....	89
Rajen Chatterjee, Gebremedhen Gebremelak, Matteo Negri and Marco Turchi Online Automatic Post-Editing across Domains.....	94
Anna Corazza, Valerio Maggio and Giuseppe Scanniello A new dataset for source code comment coherence	100
Elisa Corino and Claudio Russo Parsing di corpora di apprendenti di italiano: un primo studio su VALICO	105
Danilo Croce, Simone Filice and Roberto Basili Nystrom Methods for Efficient Kernel-Based Methods for Community Question Answering.....	111

Marco Del Tredici, Malvina Nissim and Andrea Zaninello Tracing metaphors in time through self-distance in vector spaces.....	117
Giorgio Maria Di Nunzio, Maria Maistro and Daniel Zilio Gamification for IR: The Query Aspects Game.....	123
Fabrizio Esposito, Anna Corazza and Francesco Cutugno Topic Modelling with Word Embeddings	129
Anna Fantini Spammare senza pietà - Corpus based analysis of English, unacclimatised verb loans in Italian and creation of a reference lexicon	135
Anna Feltracco, Elisabetta Jezek, Bernardo Magnini and Manfred Stede LICO: A Lexicon of Italian Connectives	141
Marcello Ferro, Franco Alberto Cardillo, Vito Pirrelli, Christina L. Gagné and Thomas L. Spalding Written word production and lexical self-organisation: evidence from English (pseudo)compounds..	146
Francesca Franzon, Giorgio Arcara and Chiara Zanini Lexical categories or frequency effects? A feedback from quantitative methods applied to psycholinguistic models in two studies on Italian.....	152
Jennifer-Carmen Frey, Aivars Glaznieks and Egon W. Stemle The DiDi Corpus of South Tyrolean CMC Data: A multilingual corpus of Facebook texts.....	157
Lorenzo Gregori, Alessandro Panunzi and Andrea Amelio Ravelli Linking IMAGACT ontology to BabelNet through action videos.....	162
Francesca Guglielmi, Pierpaolo Basile, Antonietta Curci and Giovanni Semeraro Sentiment Analysis: applicazione in un dominio psico-forense	168
Alberto Lavelli Comparing State-of-the-art Dependency Parsers on the Italian Stanford Dependency Treebank.....	173
Antonio Lieto, Enrico Mensa and Daniele P. Radicioni Taming Sense Sparsity: a Common-Sense Approach	179
Eleonora Litta, Marco Passarotti and Chris Culy Formatio formosa est. Building a Word Formation Lexicon for Latin	185
Felicia Logozzo Sequenze N+pN (nome comune + nome proprio): descrizione linguistica da un corpus dell'italiano .	190
Azzurra Mancuso, Maria De Martino and Alessandro Laudanna Semantic priming effects in Italian verbs recognition: the role of grammatical classes and semantic categories.....	195
Alessandro Mazzei Building a computational lexicon by using SQL.....	200
Anne-Lyse Minard, Manuela Speranza, Bernardo Magnini and Mohammed R. H. Qwaider Semantic Interpretation of Events in Live Soccer Commentaries.....	205
Johanna Monti, Federico Sangati, Francesca Chiusaroli, Martin Benjamin and Sina Mansour Emojitalianobot and EmojiWorldBot - New online tools and digital environments for translation into emoji	211
Giovanni Moretti, Rachele Sprugnoli and Sara Tonelli KD Strikes Back: from Keyphrases to Labelled Domains Using External Knowledge Sources.....	216

Franca Orletti, Felice Dell'Orletta and Rossella Iovino La leggibilità dei testi di ambito medico rivolti al paziente: il caso dei bugiardini di farmaci senza obbligo di prescrizione medica	222
Lucia C. Passaro, Alessandro Bondielli and Alessandro Lenci FB-NEWS15: A Topic-Annotated Facebook Corpus for Emotion Detection and Sentiment Analysis.	228
Marco Passarotti and Marco Budassi May the Goddess of Hope Help Us. Homonymy in Latin Lexicon and Onomasticon	233
Sandro Pezzelle, Ionut Sorodoc, Aurelie Herbelot and Raffaella Bernardi Imparare a quantificare guardando	237
Silvia Piccini, Andrea Bellandi, Giulia Benotto and Emiliano Giovannetti La Modellazione Diacronica di Risorse Termine-Ontologiche nell'Ambito delle Digital Humanities: Esperimenti su Clavius	243
Giulia Pieri, Dominique Brunato and Felice Dell'Orletta Studio sull'ordine dei costituenti nel confronto tra generi e complessità	248
Edoardo Maria Ponti, Elisabetta Jezek and Bernardo Magnini Grounding the Lexical Sets of Causative-Inchoative Verbs with Word Embedding	253
Martina A. Rodda, Marco S.G. Senaldi and Alessandro Lenci Panta Rei: Tracking Semantic Change with Distributional Semantics in Ancient Greek.....	258
Irene Russo, Simone Pisano and Claudia Soria Sardinian on Facebook: Analysing Diatopic Varieties through Translated Lexical Lists	263
Marco S. G. Senaldi, Gianluca E. Lebani and Alessandro Lenci Determining the Compositionality of Noun-Adjective Pairs with Lexical Variants and Distributional Semantics	268
Emilio Sulis, Cristina Bosco, Viviana Patti, Mirko Lai, Delia Irazu Hernandez Farias, Letizia Mencarini, Michele Mozzachiodi and Daniele Vignoli Subjective Well-Being and Social Media. A Semantically Annotated Twitter Corpus on Fertility and Parenthood	274
Fabio Tamburini (Better than) State-of-the-Art PoS-tagging for Italian Texts	280
Shiva Taslimipoor, Anna Desantis, Manuela Cherchi, Ruslan Mitkov and Johanna Monti Language resources for Italian: towards the development of a corpus of annotated Italian multiword expressions.....	285
Sara Tonelli, Alessio Palmero Aprosio and Francesca Saltori SIMPITIKI: a Simplification corpus for Italian.....	291
Erica Tusa, Felice Dell'Orletta, Simonetta Montemagni and Giulia Venturi Dieci sfumature di marcatezza sintattica: verso una nozione computazionale di complessità.....	297
Antonio Uva and Alessandro Moschitti Tree Kernels-based Discriminative Reranker for Italian Constituency Parsers	303
Andrea Vanzo, Danilo Croce, Roberto Basili and Daniele Nardi Context-aware Spoken Language Understanding for Human Robot Interaction.....	308
<i>Index of authors</i>	314

Written word production and lexical self-organisation: evidence from English (pseudo)compounds

Marcello Ferro
ILC-CNR Pisa, Italy

marcello.ferro@ilc.cnr.it

Franco Alberto Cardillo
ILC-CNR Pisa, Italy

francoalberto.cardillo@ilc.cnr.it

Vito Pirrelli
ILC-CNR Pisa, Italy

vito.pirrelli@ilc.cnr.it

Christina L. Gagné

Dept. of Psychology, University of Alberta, Canada
cgagne@ualberta.ca

Thomas L. Spalding

Dept. of Psychology, University of Alberta, Canada
spalding@ualberta.ca

Abstract

Elevation in typing latency for the initial letter of the second constituent of an English compound, relative to the latency for the final letter of the first constituent of the same compound, provides evidence that implementation of a motor plan for written compound production involves smaller constituents, in both semantically transparent and semantically opaque compounds. We investigate here the implications of this evidence for algorithmic models of lexical organisation, to show that effects of differential perception of the internal structure of compounds and pseudo-compounds can also be simulated as peripheral stages of lexical access by a self-organising connectionist architecture, even in the absence of morpho-semantic information. This complementary evidence supports a maximization-of-opportunity approach to lexical modelling, accounting for the integration of effects of pre-lexical and lexical access.

Il rallentamento nel tempo di battitura del primo carattere del secondo costituente di un composto inglese, rispetto al tempo dell'ultimo carattere del primo costituente, dimostra che l'implementazione del programma motorio per la scrittura di un composto è influenzata dai costituenti del composto stesso, siano essi semanticamente trasparenti o opachi. Il presente contributo offre un modello computazionale di questa evidenza, e ne valuta l'impatto sull'organizzazione del lessico mentale: la percezione del confine di morfema tra i due costituenti è analizzata come il risultato dell'interazione dinamica tra processi di accesso pre- e post-lessicale.

1 The evidence

A key question concerning the representation and processing of compound words has focused on whether (and, if so, how) morphological structure plays a role. The bulk of the research on this issue has come from recognition or comprehension tasks such as lexical decision or reading. However, written production provides a useful counterpart and allows researchers to examine whether morphemes are used even after a word has been accessed. One advantage of a typing task (in which the time to type each letter of a word is recorded) is that researchers can examine differences in processing difficulty at various points in the word. Previous research found an elevation in typing latency for the initial letter of the second constituent relative to the latency for the final letter of the first constituent for English (Gagné & Spalding 2014; Libben et al. 2012; Libben & Weber 2014) and German compounds (Sahel et al. 2008; Will et al. 2006). This elevation in typing latency at the morpheme boundary suggests that the system plans the output morpheme by morpheme, rather than as a whole unit, and that morphological programming is not complete when the motor system begins the output of the word (Kandel et al. 2008).

Gagné and Spalding (2016) examined the role of morphemic structure and semantic transparency on typing latency. The stimuli consisted in 200 compounds, 50 pseudo-compounds, and 250 monomorphemic words matched pairwise with the compounds and pseudo-compounds in the number of syllables and letters. The pseudo-compounds contain two words that do not function as morphemes (e.g., *carpet* contains *car* and *pet*). The compounds varied in whether the first and second constituent were semantically transparent. The items were displayed individually using a progressive demasking procedure and participants typed the word as the computer recorded the time required to type each letter.

The time to initiate the first letter was equivalent for monomorphemic and compound words. Typing times got faster across the word for both word types, but the rate of change was faster for monomorphemic words than for compound words. This difference was not observed when comparing monomorphemic words and pseudo-compounds.

For compounds, the rate of speed-up was slower when the first constituent was transparent than when it was opaque, but was unaffected by the transparency of the second constituent. The elevation in typing latency at the morpheme boundary was larger when the first constituent was transparent than when it was opaque, but was unaffected by the transparency of the second constituent. This difference is due to the final letter of the first constituent when the first constituent requiring less time to type when it was transparent than when it was opaque.

The data for the pseudo-compounds indicated that embedded morphemes influence production, even when they do not function as morphemes. Typing latency increased one letter prior to the end of the first constituent of a pseudo-compound and remained elevated through the boundary (e.g., both *r* and *c* in *scarcity* were elevated relative to the *a*).

1.1 Implications for lexical architectures

The reported evidence clearly indicates that morphemic structure is involved in written word production. The production of compounds differs from that of monomorphemic words and the semantic transparency of the two constituents leads to different effects. Furthermore, embedded pseudo-morphemes appear to influence the production of pseudo-compounds, but not in the same way that the embedded morphemes affect the production of compounds.

This appears to lend only partial support to models of lexical architecture where both compounds and their constituents are represented and processed as independent access units (Figure 1). In panel *A*, following Taft & Forster (1975), access and output of compounds are mediated by their constituents (*Cs*), but extra procedures would be needed to account for the role of semantic transparency in modulating the size of elevation in typing latency at the morpheme boundary. A supralexicale account (panel *B*: Grainger & Grainger 2000, Grainger et al. 1991), where constituents are activated upon compositional interpretation of compounds, cannot capture the persistence of typing effects in semanti-

cally opaque compounds (and, to an extent, in pseudo-compounds). Race models (panel *C*: Schreuder & Baayen 1995) posit parallel pathways for compound processing (both holistic and compositional), depending on variables such as whole word vs. constituent frequency, but it is not clear how they can account for effects of interaction between the two paths. Connectionist models (panel *D*: Rumelhart & McClelland 1986, Plaut & Gonnerman 2000), on the other hand, tend to dispense with specialized representational levels and access procedures, and make room for distributed effects of sublexical co-activation through overlaying patterns of processing units. A defining feature of these models is that they blur the traditional distinction between representations and processing units. We suggest that blurring this distinction can go a long way in addressing some of the issues that appear to elude models *A*, *B* and *C*.

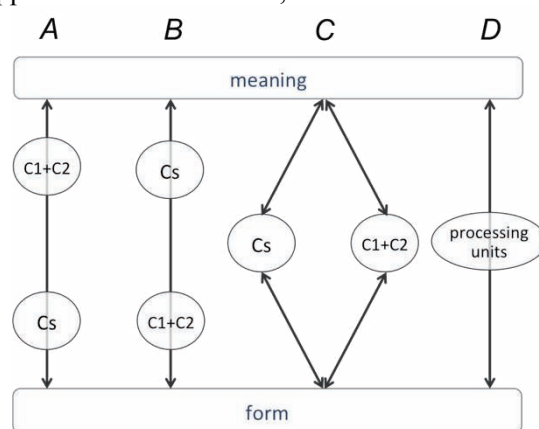


Figure 1 – Four architectures of form-meaning mapping in the mental lexicon: *C1+C2* designates two-word compounds and *Cs* mono-morphemic constituents (adapted from Diependaele et al. 2012)

Temporal Self-Organising Maps (TSOMs: Ferro et al. 2011; Marzi et al. 2014; Pirrelli et al. 2015), are a time-sensitive variant of Kohonen's SOMs (Kohonen, 2002), where words are stored through routinized, time-bound patterns of repeatedly successful processing units. Since all input words are stored concurrently on the same layer of fully connected nodes, TSOMs account for effects of co-activation of competing representations in terms of a continuous function of distributional regularities in the input data. In what follows, starting from Gagné & Spalding's evidence, we will focus on peripheral stages of lexical access/output, to verify if mechanisms of parallel, distributed pattern activation can account for differential processing effects between compounds and pseudo-compounds even in the

absence of morpho-semantic information. Although computational testing is carried out on TSOMs only, our discussion and concluding remarks address issues that go beyond a specific computational framework.

2 TSOMs

A TSOM consists of a grid of memory nodes with two layers of connectivity. The first layer (or I-layer) fully connects each node to the input vector, where symbols are sampled at discrete time ticks as patterns of activation ranging in the $[0, 1]$ interval. Weights on the I-layer are adjusted in training for individual nodes to develop specialised sensitivity to particular input symbols. Each node is also connected to all other nodes through a layer of re-entrant connections (or T-layer), whose weight strength determines the amount of influence that activation of one node has on other nodes at a one-tick delay.

When an input symbol is presented at time t , the level of activation $y_i(t)$ of node i is a function of: (a) the node’s sensitivity to the current input symbol ($y_{I_layer,i}(t)$), and (b) the re-entrant support the node receives from the map activation state at $t-1$ ($y_{T_layer,i}(t) = f(y_j(t-1))$, where f is a linear function and j ranges over all map nodes). More formally:

$$y_i(t) = \alpha \cdot y_{I_layer,i}(t) + (1 - \alpha) \cdot y_{T_layer,i}(t)$$

The node responding most strongly to the input symbol S at time tick t is called Best Matching Unit (hereafter $BMU(S, t)$ or $BMU(t)$ for short).

The map’s response to a sequence of input symbols like *carpet* is a chain of consecutively firing $BMUs$, each responding to a letter in *carpet*. During training, connection weights between consecutive $BMUs$ are adjusted to the frequency distribution of input symbols in the training set, according to Hebbian principles of correlative learning. Given the bigram ab , the connection strength between $BMU(a, t-1)$ and $BMU(b, t)$ increases if a often precedes b (entrenchment) and decreases if b is often preceded by a symbol other than a (competition) (Figure 2, left). Combination of entrenchment and competition yields selective specialisation of chains of $BMUs$ (Figure 2, right). If the same input symbol follows different contexts, it will tend to be responded to by more $BMUs$, one for each context. The stronger the probabilistic support that the input symbol receives from its preceding context, the more likely the recruitment of a dedicated BMU ,

and the stronger its re-entrant connection. As a result of this dynamic, high-frequency words recruit specialised node chains, low-frequency words are responded to by weaker, “blended” node chains.

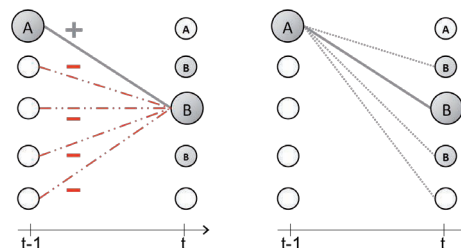


Figure 2 - Left: operation of Hebbian rules on potentiated (+) and inhibited (-) connections. Right: forward one-tick-delay connections leaving ‘A’ at time $t-1$. Larger nodes represent $BMUs$. Shades of grey indicate levels of node activation.

2.1 The Experiment

The 200 compounds and 50 pseudo-compounds used by Gagné & Spalding were used to train a 40×40 node TSOM for 100 learning epochs. Besides compounds and pseudo-compounds, the training set included 500 (pseudo)constituents as individual words (e.g. *car* and *wash* in *carwash*, *car* and *pet* in *carpet*), for a total amount of 750 items. At each training epoch, monomorphemic words were shown 10 times as often as compounds. We ran 5 repetitions of the experiment, and results were analysed using linear mixed effects models (LME), with experiment repetitions and training items as random variables.

To analyse differential processing effects for pseudo-compounds and compounds, we focused on two types of evidence: (i) per-letter performance of a trained TSOMs in incrementally anticipating compounds and pseudo-compounds; (ii) structural connectivity of $BMUs$ responding to letter bigrams at the C1-C2 boundary.

To anticipate a progressively presented input word, a TSOM propagates the activation of the current $BMU(t)$ through its forward temporal connections, and outputs, at each time tick, the symbol $S_{BMU(t+1)}$ encoded on the I_layer of the most strongly (pre)activated node:

$$BMU(t+1) = \operatorname{argmax}_{i=1,\dots,N} \{m_{i,h}\} \quad h = BMU(t)$$

where $m_{i,h}$ is the weight value on the forward temporal connection from node h to node i . Each correctly predicted symbol in the input word is assigned the prediction score of the preceding symbol incremented by 1. Otherwise, the symbol receives a 0-point score.

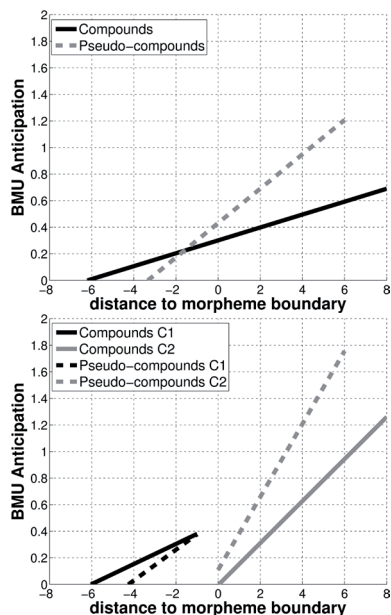


Figure 3 – Marginal plots of interaction effects between compounds vs. pseudo-compounds and letter distance to morpheme boundary in an LME model fitting anticipation of up-coming *BMUs* by a TSOM. Negative and positive x values indicate letter positions located, respectively, in the first and second constituent. Anticipation is plotted across whole (pseudo)compounds (top panel), and by individual constituents (bottom panel).

Figure 3 (top panel) illustrates the rate of letter anticipation across the word for both compounds and pseudo-compounds, plotted by distance to the morpheme boundary. The steeper rate for pseudo-compounds than for compounds shows that pseudo-compounds are easier to predict/anticipate than compounds. We take this evidence to be in line with evidence of a faster speedup rate in the typing of monomorphemic vs. compound words. A closer look at anticipation rates for individual constituents (Figure 3 bottom panel) shows a drop of anticipation at the C1-C2 boundary (more prominent for compounds than pseudo-compounds) with a steeper increase in C1 and C2 for pseudo-compounds, which happen to be, on average, shorter than C1 and C2 in real compounds.

To look for structural correlates of anticipation rates in the map, we conducted, for each item, a letter-by-letter analysis of values of pointwise entropy (PWH) for the connections between consecutive *BMUs*, namely $h=BMU(t-1)$ and $i=BMU(t)$:

$$PWH(m_{i,h}) = -\log \frac{m_{i,h}}{\sum_k m_{k,h}}$$

The value of PWH for the connection between end-C1 and start-C2 ($x = 0$) has a local peak in compounds only (Figure 4). Since PWH provides a measure of how unexpected the activation of

$BMU(t)$ is, this structural evidence can account for a delay in processing and a drop in anticipation at the morpheme boundary of compounds, but not of pseudo-compounds.

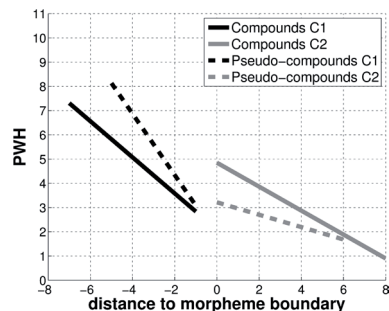


Figure 4 – Marginal plots of interaction effects between compound vs. pseudo-compound constituents and letter distance to morpheme boundary in an LME model fitting pointwise entropy of forward *BMU* connections. Negative and positive x values indicate letter positions located, respectively, in the first and second constituent.

3 Discussion and conclusions

Trained on both compounds and pseudo-compounds, TSOMs develop a growing sensitivity to surface distributional properties of input data, turning chains of randomly connected, general-purpose nodes into specialised sub-chains of *BMUs* that respond to specific letter strings at specific positions. Compounds not only tend to occur, on average, less frequently than their C1/C2 constituents do as independent words (Ji et al. 2011), but they tend to present lower-frequency bigrams at the C1-C2 boundary than do pseudo-compounds. Principles of Hebbian learning allow TSOMs to capitalise on both effects. Entrenchment makes expectations for high-frequency bigrams stronger and expectations for low-frequency bigrams weaker. At the same time, the competition between C1 as an independent word and C1 as the first constituent in a C1-C2 compound biases the map's expectation towards the most frequent event (C1 in isolation). Compound families, i.e. sets of compounds sharing C1 (*windmill*, *windshield* etc.) or C2 (*snowball*, *basketball* etc.), magnify these effects, making the map more sensitive to formal discontinuity at morpheme boundaries. When more C2s can follow the same C1 in complementary distribution, the left-to-right expectation for a particular C2 to occur, given C1, decreases. Likewise, when more C1s competitively select the same C2, the individual contribution of each C1 to the prediction of C2 decreases. We conjecture that more global effects of lexical organisation like these may eventually blur local memory

effects based on position-independent bigram frequencies.

Our simulations with TSOMs can model the correlation between continuously varying distributional regularities in the input data and peripheral levels of routinized recognition and production patterns. These patterns are in line with Gagné & Spalding's evidence of (a) the influence of embedded pseudo-morphemes on cascaded models of written word production, and (b) faster anticipation rates for monomorphemic vs. compound words.

Further experimental results (not reported here), obtained by including compound families in the training data, confirm slower anticipation rates for true compound constituents, due to the combined effect of word frequency distributions and word compositionality in compound families. The size of a compound family can arguably be a function of the degree of productivity and semantic transparency of its members (Baroni et al. 2007). The influence of the compound family size on anticipation rates can shed light on the influence of levels of semantic transparency on compound processing. Simulation evidence suggests that the bigger the family, the stronger its influence will be. Finally, we also monitored the influence of increasing token frequencies of monomorphemic words in the training data on the map perception of constituent boundaries within compounds. As expected, for constant frequency values of compounds in the training set, the higher the token frequency of monomorphemic words, the higher the pointwise entropy of connections at the C1-C2 boundary.

A full account of Gagné & Spalding's evidence of a graded influence of semantic transparency on compound processing is beyond the reach of the computational architecture presented here. Surface effects of discontinuity in the internal structure of compounds (as opposed to pseudo-compounds) appear to provide a purely formal, pre-lexical scaffolding for truly morpho-semantic effects to emerge at later processing stages. To model these effects, we appear to be in need of a parallel processing architecture able to effectively integrate several representational levels (orthographic, phonological, morphological, and conceptual) and different processing steps within a single distributed system (Smolka et al. 2009). Nonetheless, our simulations show that by letting compounds, pseudo-compounds and (pseudo)constituents compete for the same level of memory resources on a topological map, it is possible to account for apparently contradic-

tory effects of a) graded perception of constituent boundary in both compounds and pseudo-compounds, apparently requiring prelexical decomposition, and b) higher anticipation rates for pseudo-compounds than compounds, supporting full form representations for lexical access.

References

- Baroni M., Guevara E., & Pirrelli V. (2007) NN Compounds in Italian: Modelling Category Induction and Analogical Extension, in V. Pirrelli (Ed.) *Psycho-computational issues in morphology learning and processing. Lingue e Linguaggio VI* (2), 263 - 290.
- Diependaele, K., Grainger, J., & Sandra, D. (2012). Derivational Morphology and Skilled Reading: An Empirical Overview. In M. J. Spivey, J. McRae, & M. F. Joanisse (Eds.) *The Cambridge Handbook of Psycholinguistics*, 311-332. Cambridge University Press.
- Ferro M., Marzi, C., & Pirrelli, V. (2011). A Self-Organizing Model of Word Storage and Processing: Implications for Morphology Learning. *Lingue e Linguaggio X* (2), 209 - 226.
- Gagné, C. L., & Spalding, T. L. (2014). Typing time as an index of morphological and semantic effects during English compound processing. *Lingue e Linguaggio, XIII*(2), 241-262.
- Gagné, C. L., & Spalding, T. L. (2016). Effects of morphology and semantic transparency on typing latencies in English compound and pseudo-compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 1489-1495.
- Giraud, H., & Grainger, J. (2000). Effects of prime word frequency and cumulative root frequency in masked morphological priming. *Language and Cognitive Processes*, 15(4/5), 421-44.
- Grainger, J., Colé, P., & Segui, J. (1991). Masked morphological priming in visual word recognition. *Journal of Memory and Language*, 30, 370-84.
- Kandel, S., Álvarez, C. J., & Vallée, N. (2008). Morphemes also serve as processing units in handwriting production. In M. Baciú (Ed.), *Neuropsychology and cognition of language. Behavioural, neuropsychological and neuroimaging studies of spoken and written language*, 87-100. Kerala, India: Research Signpost.
- Libben, G. (2010). Compound words, semantic transparency, and morphological transcendence. *Linguistische Berichte, Sonderheft 17*, 317-330.
- Libben, G., & Weber, S. (2014). Semantic transparency, compounding, and the nature of independent variables. In F. Rainer, W. Dressler, F. Gardani, &

- H. C. Luschutzky (Eds.), *Morphology and meaning*. Amsterdam: Benjamins.
- Libben, G., Weber, S., & Miwa, K. (2012). P3: A technique for the study of perception, production, and participant properties. *The Mental Lexicon*, 7(2), 237-248. doi:10.1075/ml.7.2.05lib
- Marzi, C., Ferro, M., & Pirrelli, V. (2014) Morphological structure through lexical parsability. *Lingue e Linguaggio XIII (2)*, 263-290.
- Pirrelli, V., Ferro, M., & Marzi, C. (2015). Computational complexity of abstractive morphology, In Baerman M., Brown D., Corbett G. (Eds.) *Understanding and Measuring Morphological Complexity*. 141-166, Oxford, United Kingdom: Oxford University Press.
- Plaut, D. C., & Gonnerman, L. M. (2000). Are nonsemantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, 15(4/5), 445–85.
- Sahel, S., Nottbusch, G., Grimm, A., & Weingarten, R. (2008). Written production of German compounds: Effects of lexical frequency and semantic transparency. *Written Language & Literacy*, 11(2), 211-227. doi:10.1075/wll.11.2.06sah
- Schreuder, R., & Baayen, H.R. (1995) Modeling morphological processing. In Feldman, L. B. (Ed.) *Morphological aspects of language processing*, 131–56, Hillsdale, NJ: Erlbaum.
- Smolka, E., Komlosi, S., & Rösler, F. (2009). When semantics means less than morphology: The processing of German prefixed verbs. *Language and Cognitive Processes*, 24(3), 337-375.
- Rumelhart, D., & McClelland, J. 1986. On learning the past tense of English verbs. In Rumelhart, D.E, McClelland J. (Eds.), *Parallel distributed processing: Explanations in the microstructure of cognition*, 216-271, The MIT Press.
- Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14, 638–47.
- Will, U., Nottbusch, G., & Weingarten, R. (2006). Linguistic units in word typing: Effects of word presentation modes and typing delay. *Written Language & Literacy*, 9(1), 153-176.