1st Joint Conference of the EPS and SEPEX, Granada 2010

Monitoring the penalization/advantage of lexical ambiguity in vector model representations

Guillermo Jorge-Botana, Ricardo Olmos & José A. León
Universidad Autónoma de Madrid

Extensional definition of lexical ambiguity

- Two criteria for categorizing the phenomenon
  - Representation of meanings

    Separate *vs.* single representation of senses

    (Homonymy, Polysemy, Metonymy, Metaphors…)
  - Contextual distribution of word occurrences

    Occurrences across few or many contexts

Representation of senses

- Separate senses (Klein & Murphy, 2001).
- A common core with specific parts

  (Rodd, Gaskell & Marslen-Wilson, 2002; Klepousniotou & Baum, 2006).
- Unique representation as in LSA models (Kintsch, 2001; Kintsch & Mangalath, in press).

Contextual distribution of word occurrences

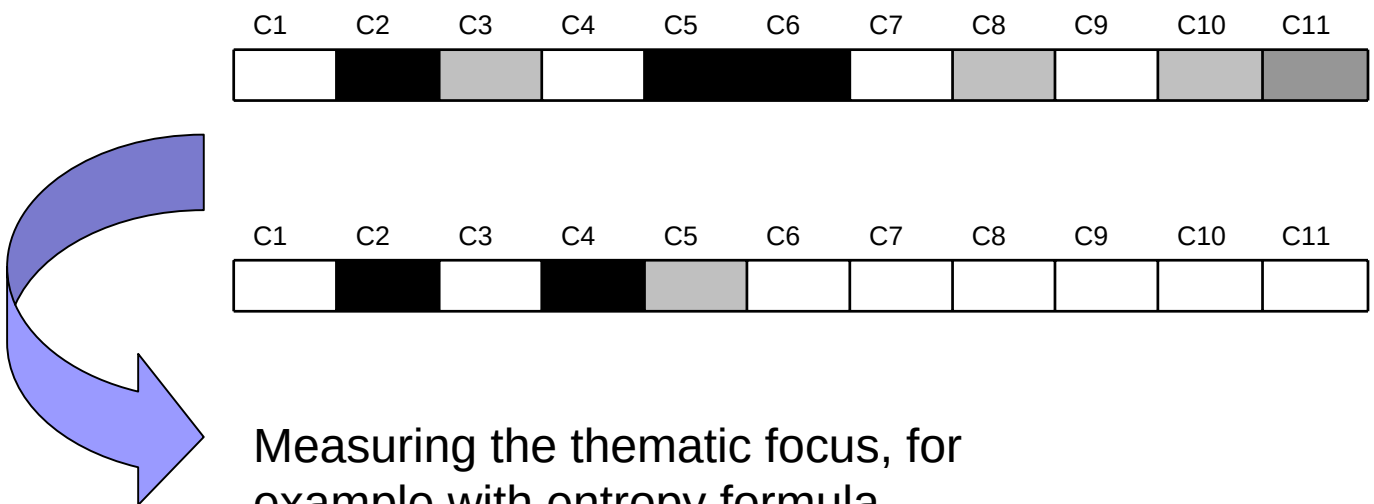
- Measured by Contextual Diversity (Adelman, Gordon, Brown & Quesada, 2006, McDonald & Shillcock, 2001)
  - **Polysemy**: Polysemic words have greater Contextual Diversity than monosemic.

Since Abstract words have greater Contextual Diversity than Concrete words …

  - Can **abstractness** be a kind of lexical ambiguity?

Concept of Context Diversity in Vector-Models

The vectors that represent the words inside a Vector-Space (as in LSA) can be a measure of indices such Context Diversity



Current Study

- So what can a model such as LSA tell us about some of the empirical effects related to lexical ambiguity?
  - A model of a unique lexical representation
  - An exclusively linguistic model.

Current Study

- Effects related to lexical ambiguity

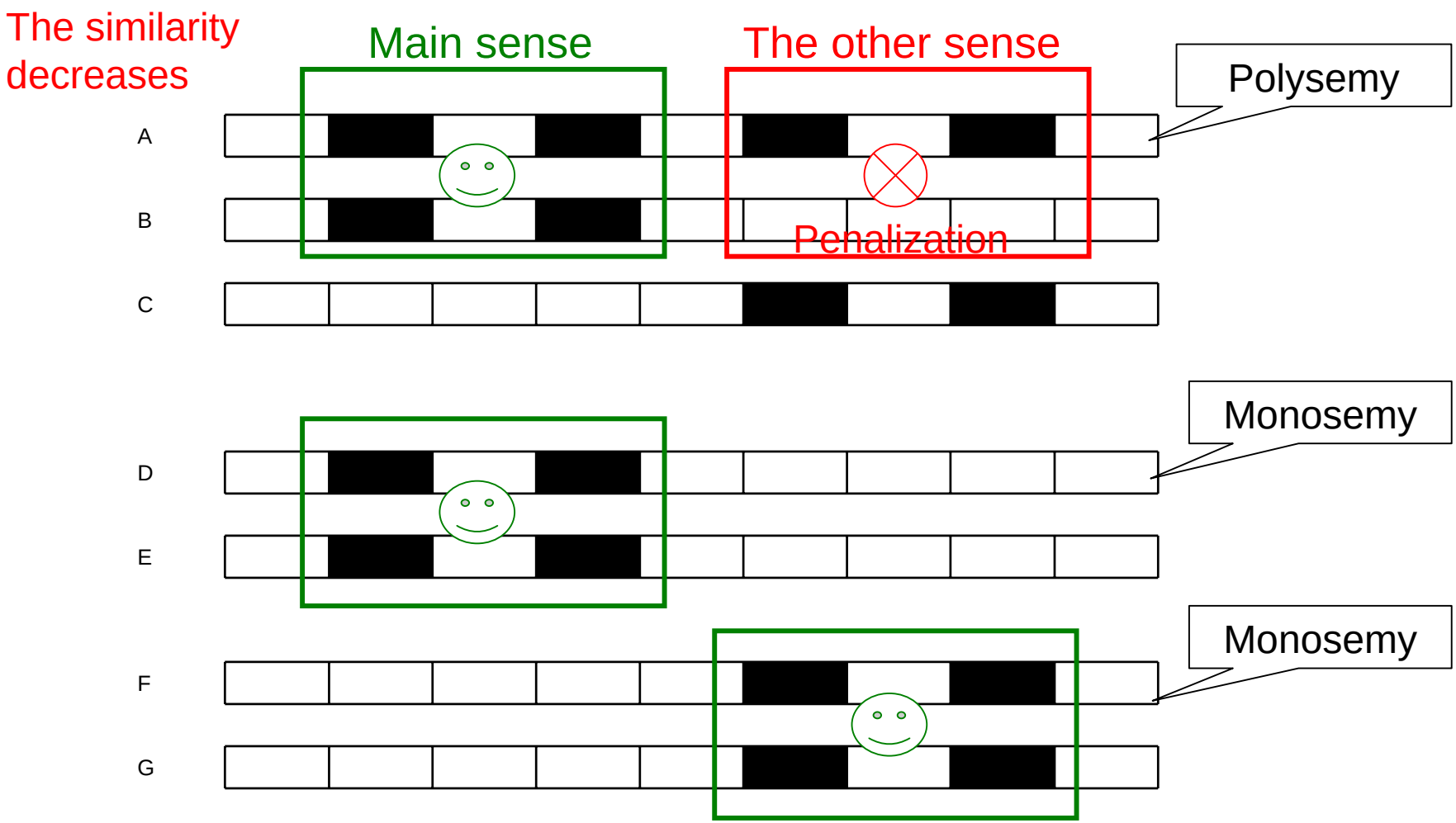
1. **Associative difficulty**
2. **Lexical Decision Task**

1. Associative difficulty

- For polysemic words: difficulty generating meaning without a context (Duffy, Morris, & Rayner, 1988)

- Abstract words rarely have semantic relations (they predominantly have associative relations)
  (e.g. Crutch, 2006; Crutch, Ridha, & Warrington, 2006; Crutch & Warrington, 2005; Warrington & Crutch, 2007; Crutch & Warrington, 2005; Duñabeitia et al., 2009)

Associative difficulty for LSA



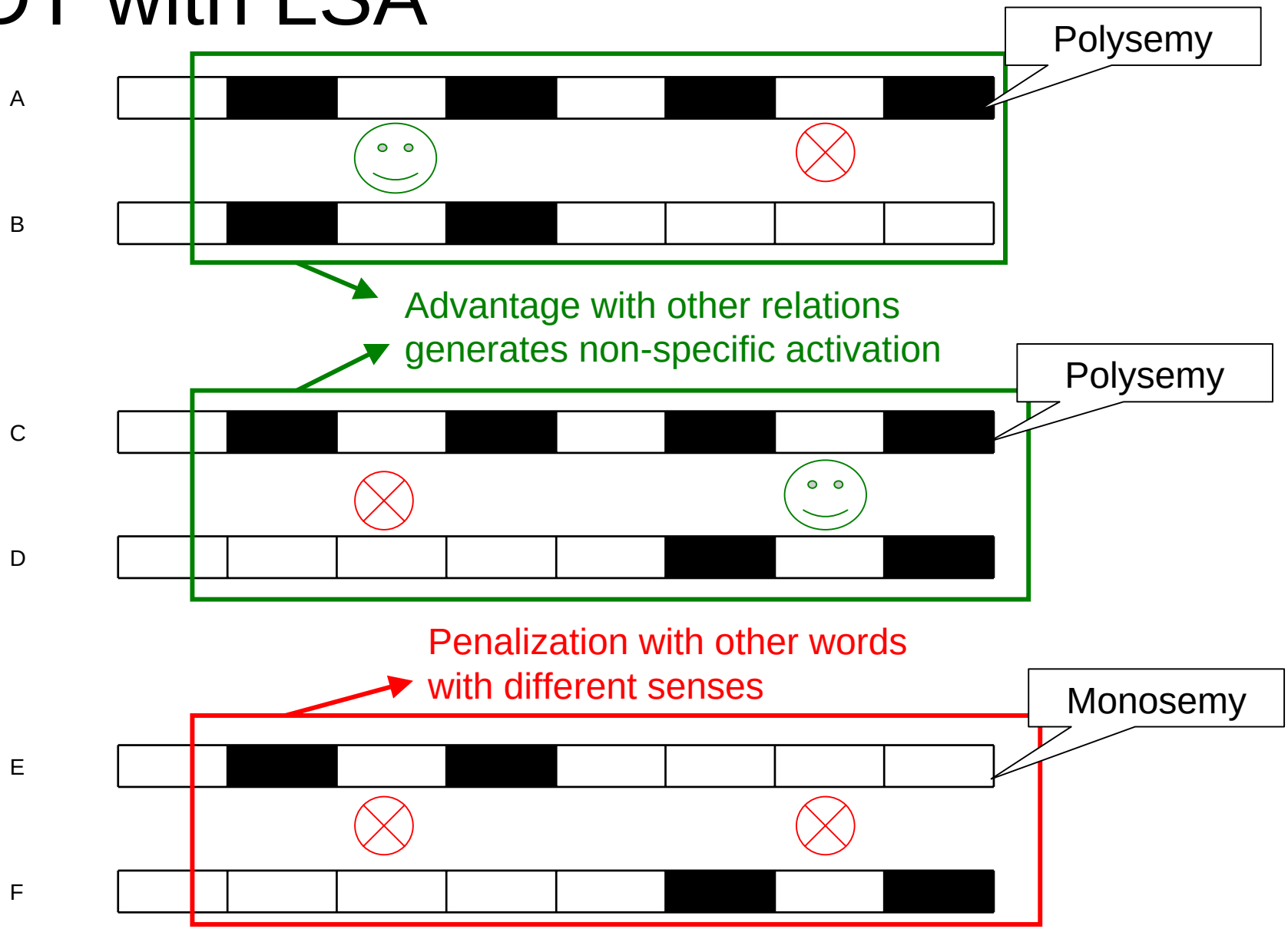
2. Data with LDT

- Polysemy advantage (Hino & Lupker, 1996; Pexman & Lupker, 1999)
- Concrete advantage
- The response to LDT is task dependent
- It is favored by unspecific activation mediated by semantic representations (Piercey & Joordens, 2000)

LDT with LSA

- Ambiguous vectors have difficulty with main relations but an advantage with other relations.
- This is due to the distributional properties of the vectors.
- This advantage with other relations can produce non-specific activation, improving LDT response times

LDT with LSA


Polysemy
A
B
Advantage with other relations
generates non-specific activation
Polysemy
C
D
Penalization with other words
with different senses
Monosemy
E
F

Hypotheses

- **A vector model like LSA can simulate the associative difficulty**
  - In polysemic words
  - In abstract words
- **A vector model can simulate the response to LDT in polysemic words**
- **Since responses to LDT are the opposite in Abstract words (to polysemic). A linguistic model like LSA cannot simulate the response of abstract words to LDT without introducing another source of activation (for example mental imagery).**

Procedure

- Three simulations
  1. Introducing an invented word into semantic space at different points along the polysemy-monosemy continuum.
  2. With real polysemic-monosemic words
  3. With real abstract-concrete words

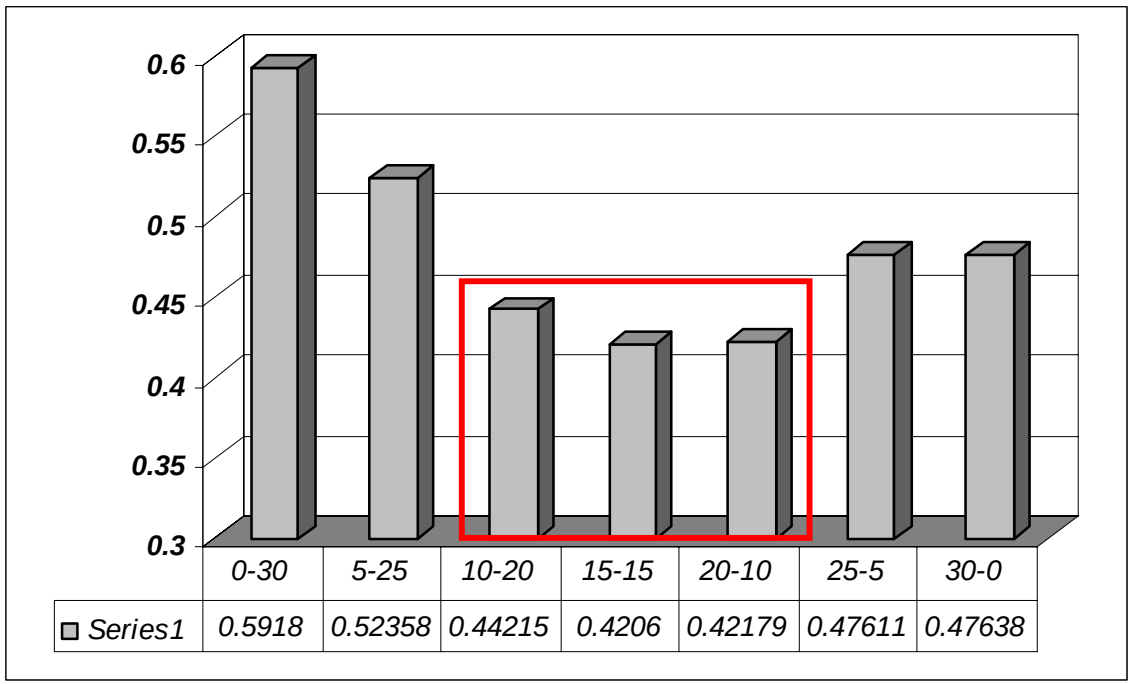
The analysis

- Extract semantic neighbors and analyze the similarities (cosines) between them and the reference word.
- Analyze Context Diversity within the vector with ranges and entropy measures.

First simulation

- LSA is trained with LEXESP corpus (Sebastián, Cuetos, Carreiras & Martí, 2000)
- An artificial nonsense word "**noray**" is introduced into the space in documents with two possible senses represented in different proportions: the sense of "traditional sport" and sense of "old neighborhood"
- The proportions of the two senses were
  - 30-0 (monosemy),
  - 25-5 (dominant polysemy)
  - 15-15 (equally probable polysemy)
  - 5-25 (dominant polysemy)
  - 0-30 (monosemy)

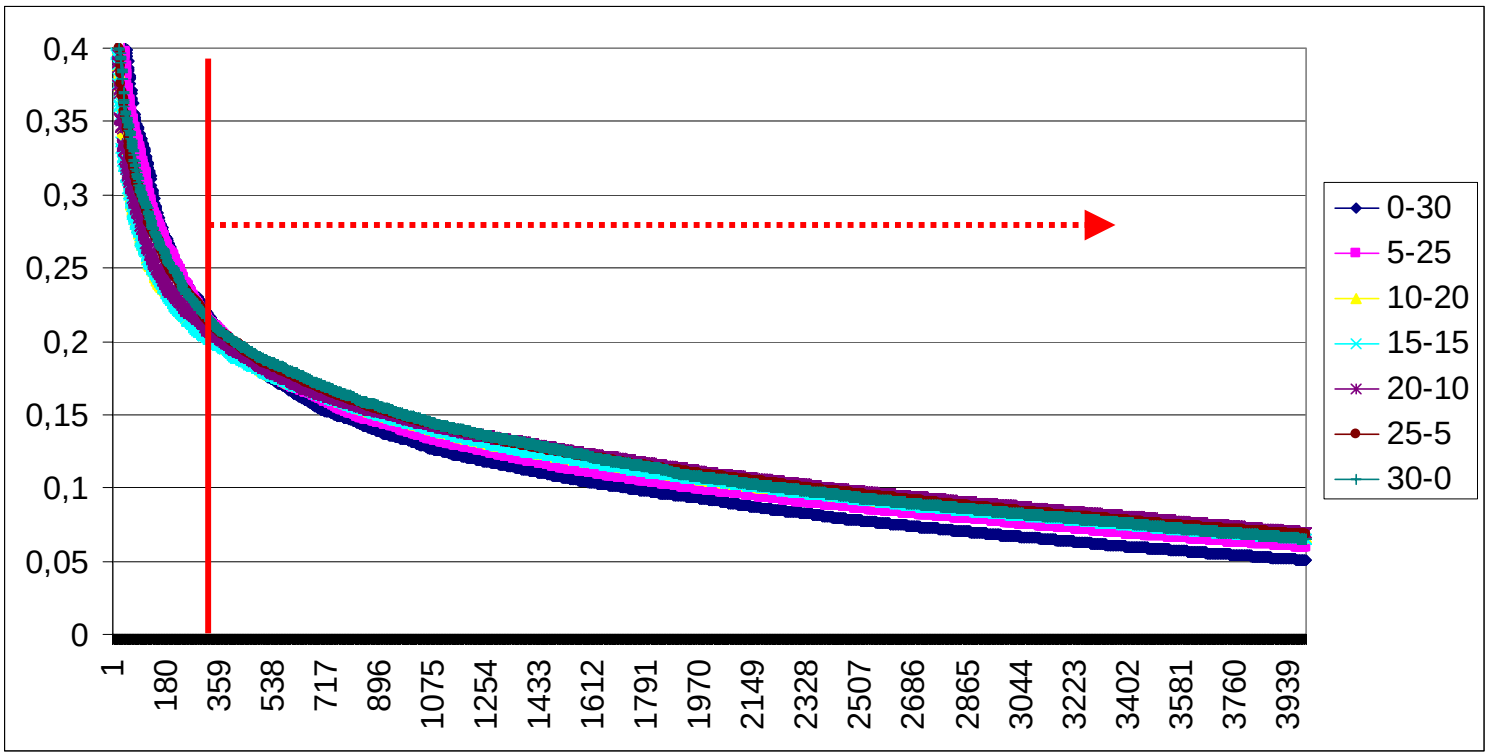
Results(I)



Mean of the 30 first semantic neighbors

The penalization exists for polysemic conditions

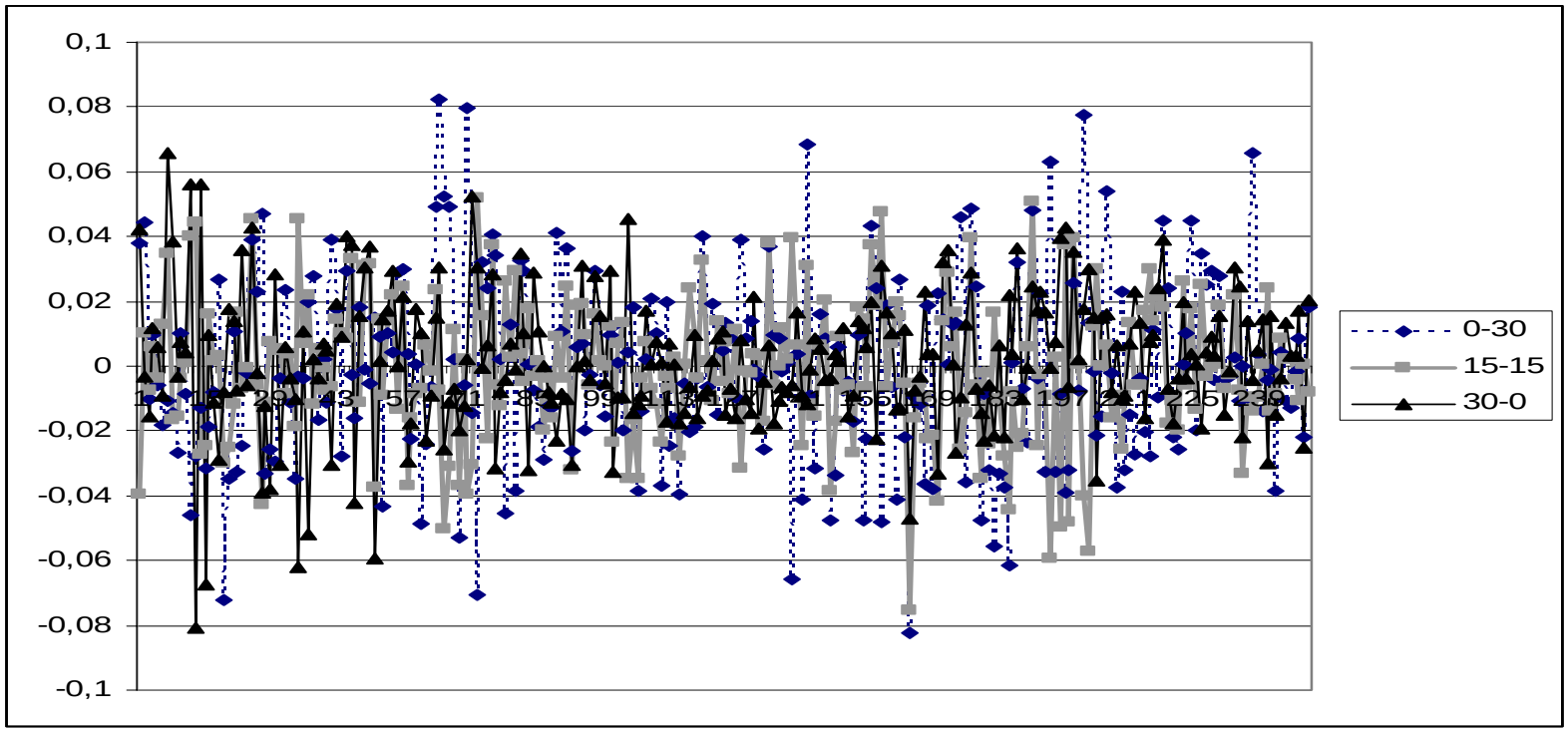
Results(I)



Ranges of the 4000 first semantic neighbors

- The pattern changed around neighbor 350 (cosine 0.2).
- For the polysemic conditions, the penalization became advantageous from this point.

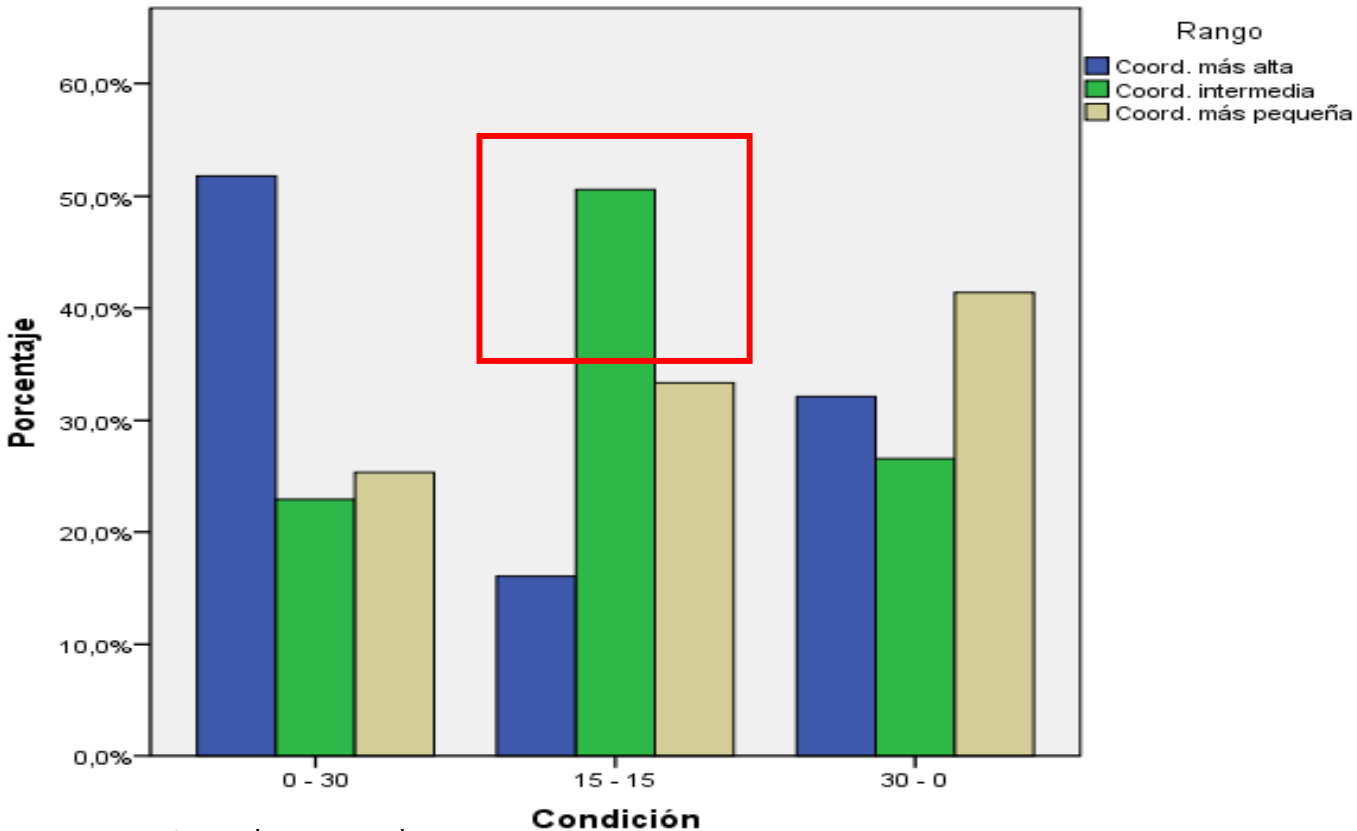
Results(I)



Scores on the dimension of the vectors in the three main conditions

- The effect of the penalization and the change of pattern may be due to the distributional properties of the vectors.
- The 15-15 condition was spread along all dimensions.

Results(I)



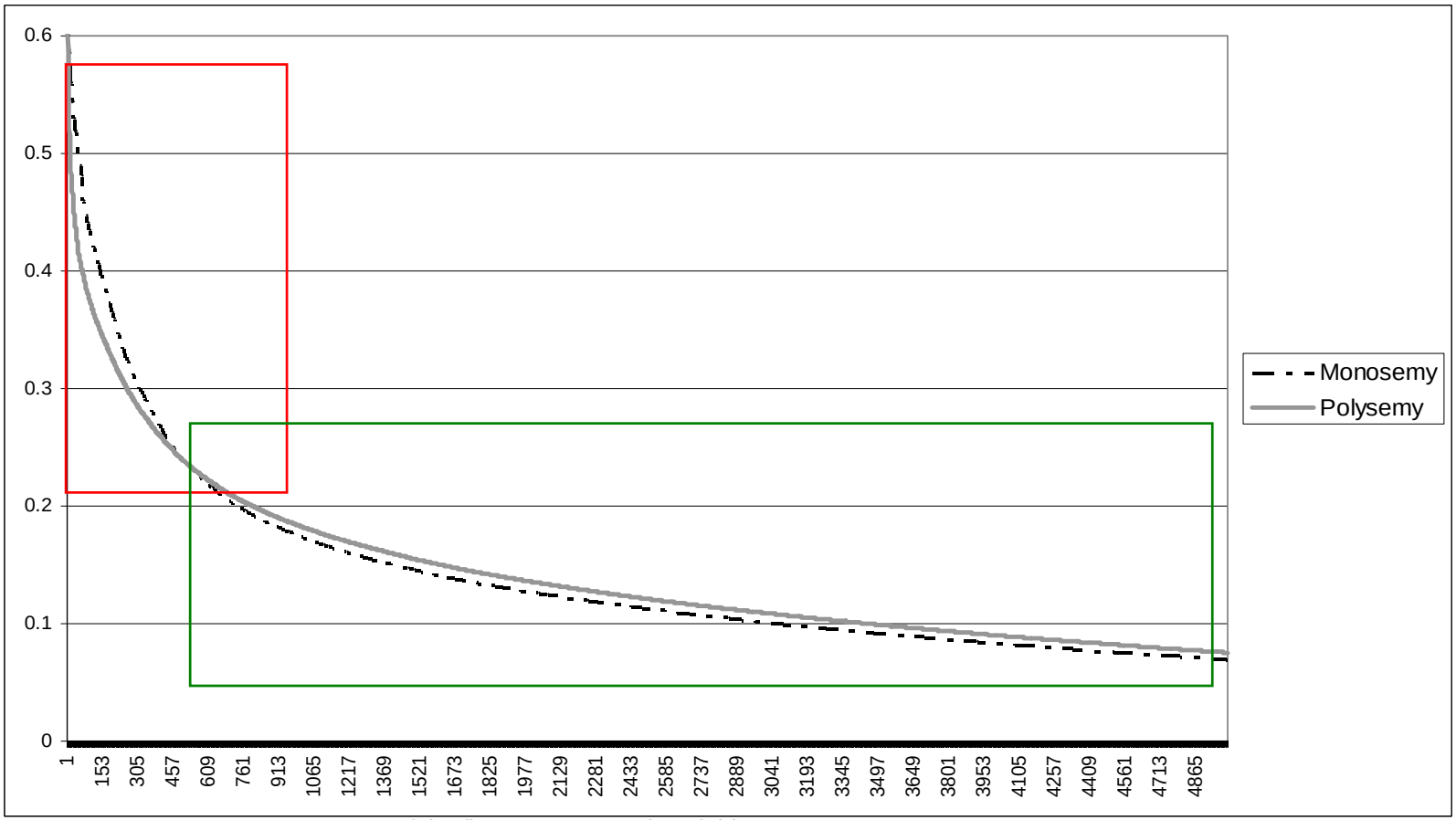
1st, 2nd and 3rd position in the saturations of dimensions

- The position of polysemic condition (15-15) were often medium (scoring low on all the dimensions).

Second simulation

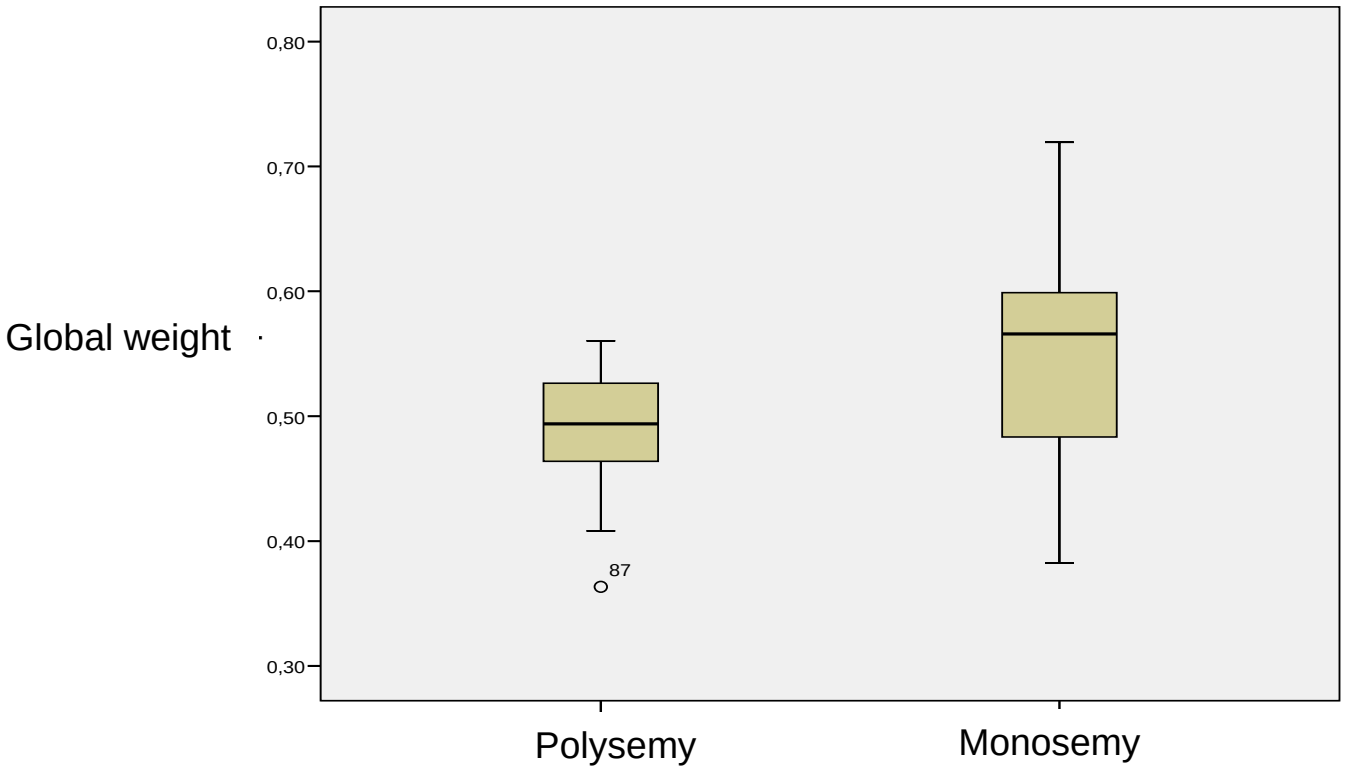
- Same as first simulation but with real **polysemic and monosemic** words extracted from other studies (Estévez, 1991; Jorge-Botana, León & Olmos, 2010).
- Controlled for frequency, concreteness and imaginability.

Results (II)



- Monosemy had an advantage for the first relations.
- The pattern changes around neighbor 550 (cosine 0.2) in favour of polysemic.

Results(II)

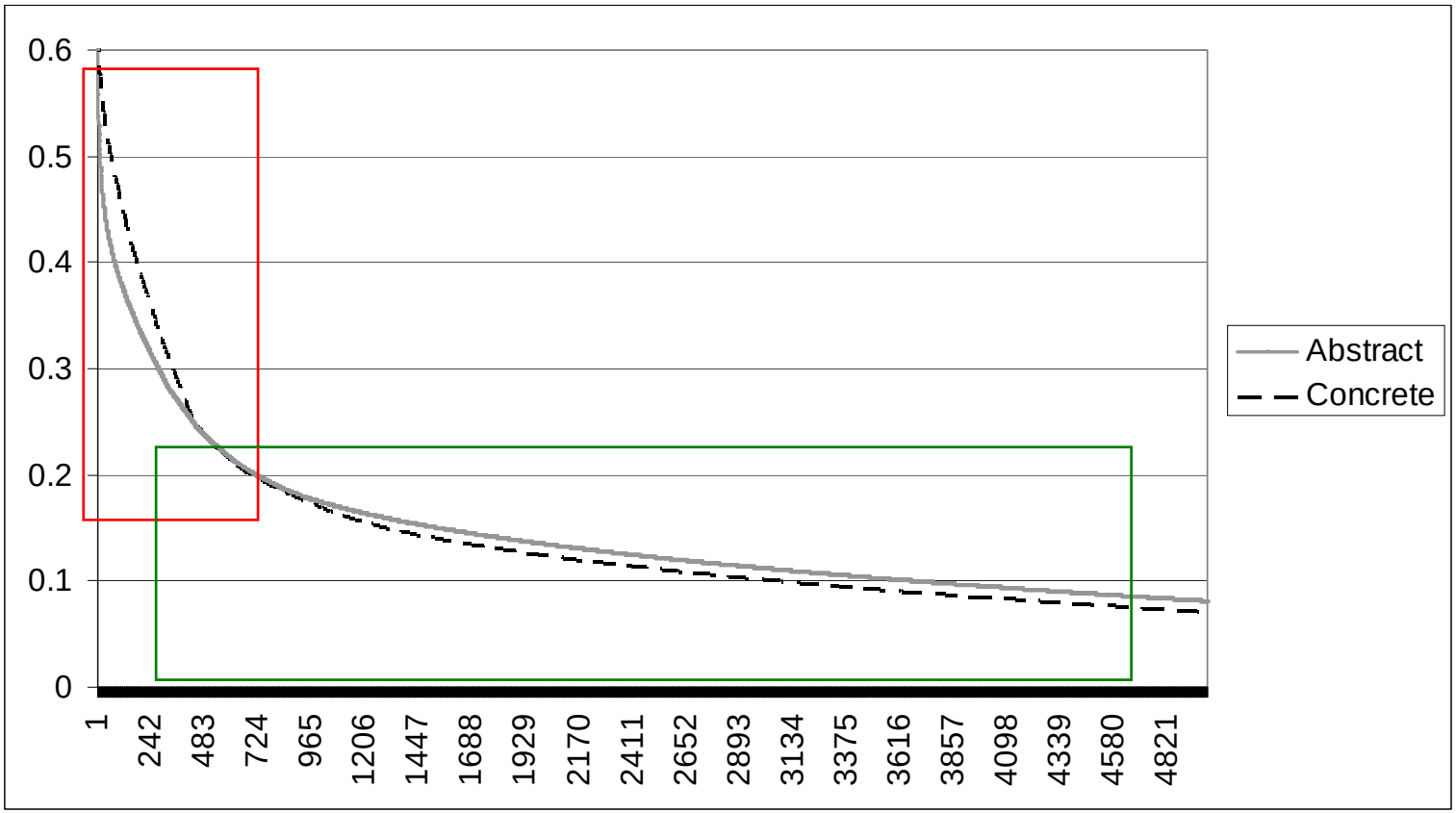


- The polysemy condition have less global weight thus more entropy
- The spread among different contexts for polysemic words is a substantive fact.

Third simulation

- Same as second simulation but with real **abstract and concrete** words extracted from other studies (Duñabeitia et al., 2009).

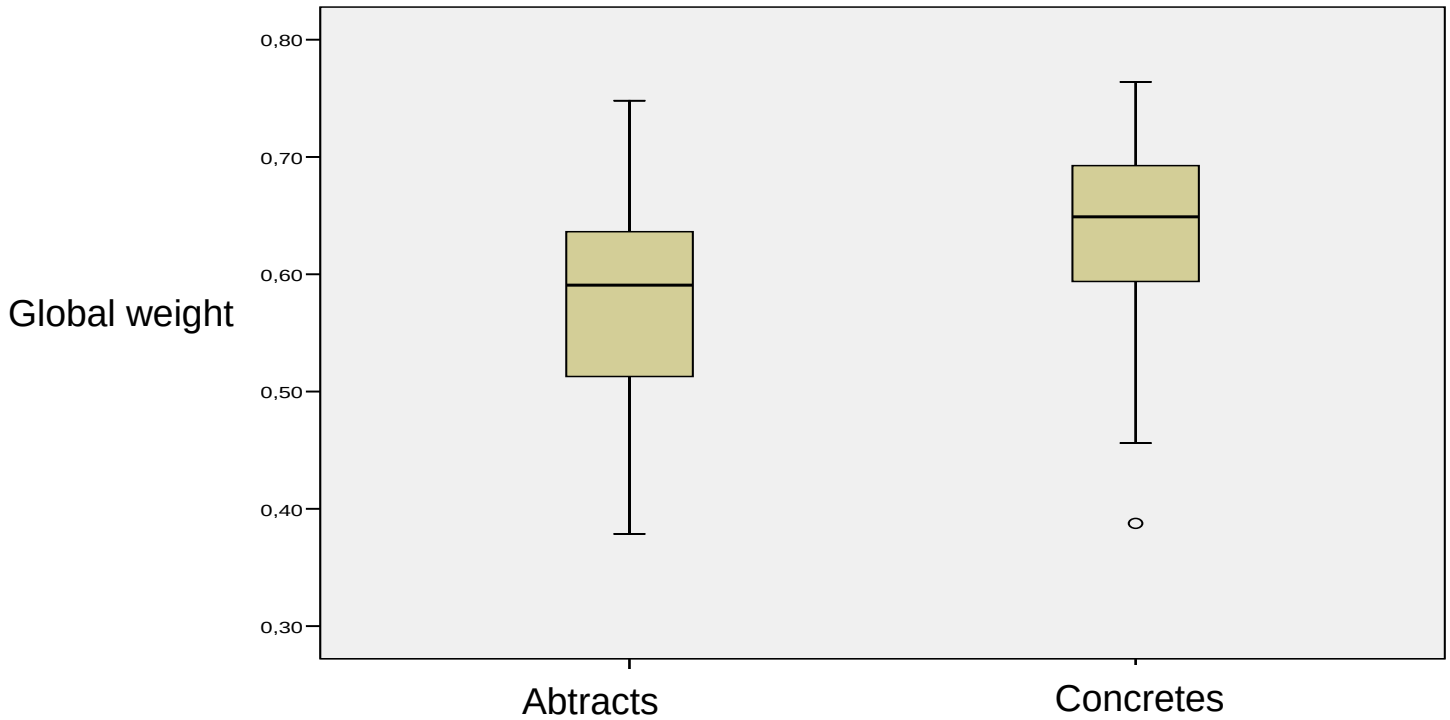
Results(III)



Ranges of the first 5000 semantic neighbors

- Concrete words obtained an advantage for the first 400 relations.
- The pattern changed around neighbor 400 (cosine 0.2) in favor of abstract words.

Results(III)



- The abstract condition had less global weight thus more entropy.
- The spread among different contexts for abstract words is a substantive fact.

Preliminary conclusions

- So, what can a model like LSA tell us about the empirical effects in question?
  - A model of a unique representation
  - An exclusively linguistic model.

Preliminary conclusions (I)

In **polysemy** and in **abstractness**, associative difficulty could be explained by the linguistic distributional properties of a unique representation.

Preliminary conclusions (II)

In **polysemy**, LDT responses could be explained by the unspecific activation produced by the linguistic distributional properties of a unique representation.

Preliminary conclusions (III)

- In the **Abstract** word condition, LDT responses couldn't be explained by the linguistic distributional properties. Another source of potential activation is required.

(perceptual representations?)

Preliminary conclusions (IV)

- In abstract word condition:
  - No additional source other than linguistic is needed as an explanation to account for associative difficulty.
  - No differences in cerebral activation between concrete and abstract in semantic tasks (Pexman et al., 2007).
  - Another source of activation is needed for LDT responses.
  - Differences in cerebral activation between concrete and abstract in LDT (Binder, Westbury, et al, 2005).

Thank you

www.elsemantico.com

Monitoring the penalization/advantage of lexical ambiguity in vector model representations

Guillermo Jorge-Botana, Ricardo Olmos & José A. León

Universidad Autónoma de Madrid