A cognitive perspective on Latent Semantic Analysis

a tutorial at the First European Workshop on LSA in Technology-Enhanced Learning

Benoît Lemaire

Laboratoire TIMC-IMAG (CNRS UMR 5522)
University of Grenoble
France
Benoit.Lemaire@imag.fr
Outline

- Computational cognitive modeling
- LSA-based cognitive models
  - Semantic memory
  - Vocabulary acquisition
  - Text comprehension
- Competing models
Intuitive introduction

Corpus

LSA

flew
flying
pilot
plane
airport
truck
car
train
drive
parks
table
restaurant

bee
bird
wings

nectar
petals
bouquet
flowers

grow
flower
garden
tree
fruit

bananas
apples
vegetables

flying
honey
growing

flowers

Intuitive introduction
Intuitive introduction

texts → LSA → meaning
Intuitive introduction

texts $\rightarrow$ LSA $\rightarrow$ meaning
TEXTS ➔ MEANING cognitive processes

- Semantic memory
TEXTS ➔ MEANING cognitive processes

- Semantic memory
- Vocabulary acquisition
texts ➔ meaning cognitive processes

- Semantic memory
- Vocabulary acquisition
- Text comprehension
TEXTS ➔ MEANING cognitive processes

- Semantic memory
- Vocabulary acquisition
- Text comprehension
- Summary assessment

texts ➔ meaning
Method

texts → LSA → meaning

estimate

texts → meaning

LSA
- Study psycholinguistic theories
- Design and implement a cognitive model
- Collect a corpus
- Run the model on the corpus
- Collect experimental data
- Compare model and data
- Update the model
Models based on LSA

- Semantic memory
- Vocabulary acquisition
- Text comprehension
LSA, a good model of semantic memory?

- Comparison of words
  - TOEFL test
  - Vocabulary test for children
- Comparison of sentences
  - Measure of coherence
- Comparison of texts
Word comparisons

- Synonymy part of the TOEFL test (Landauer & Dumais, 97)
  - 80 items: 1 word/4 alternative words
  - Guess which one is the closest
  - Example: levied
    - imposed
    - believed
    - requested
    - correlated
  - LSA was trained on a 4.6 million word corpus
  - LSA score: 64.4%
  - Foreign applicants to US colleges: 64.5%
Cognitive plausibility

- LSA: 64.4%  Humans: 64.5%
- A good point for the model
- Better computational techniques:
  - Combination of corpus-based and lexicon-based methods: 97.5% (Turney, 2003)
- ... but less cognitively plausible
Vocabulary test: LSA vs children
(Dehierre & Lemaire, 2004)

- 115 items
- 4 definitions: correct, close, distant, unrelated
- Subjects were in grade 2 to 4
- Example (translated from French):

  slope?

  - rising road
  - tilted surface which goes up or down
  - small piece of ground
  - face of a rock or a mountain
A children's semantic space

- Stories and tales for children: ~ 1,600,000 words
- Children productions: ~ 800,000 words
- Textbooks: ~ 400,000 words
- Encyclopedia: ~ 400,000 words
- Dictionary: ~ 50,000 words

TOTAL: 3,2 million words
Vocabulary test: results

Children data
Vocabulary test: results

Children + Model data

Correct | Close | Far | Unrelated

2nd grade
3rd grade
4th grade
LSA
99 most frequent words

Children+Model data

Correct  Close  Far  Unrelated

2nd grade  3rd grade  4th grade  115 words  99 words
Vocabulary test: word lemmatization

Children+Model data

Correct | Close | Far | Unrelated
---|---|---|---
2nd grade | 3rd grade | 4th grade | 115 words | 99 words | lemm.
0 | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70
Vocabulary test: verb lemmatization

Children+Model data

Correct Close Far Unrelated

2nd grade 3rd grade 4th grade

115 words 99 words lemm. verb lemm.
Other experiments show no improvement of stemming on LSA similarities

Possible explanation:
- Various forms may not have the same meaning. They do not appear in the same contexts
Comparison of sentences

- Use LSA to measure the coherence of a text
- Compute the semantic similarities between adjacent sentences

\[
\text{coherence (T)} = \frac{\sum_{i=1}^{n-1} \cosine(\text{sentence } i, \text{ sentence } i+1)}{n-1}
\]
Foltz et al. (1998) experiment

- McNamara et al. (1996) data
- 4 versions of text about the heart
  - Low local coherence, low macro coherence (lm)
  - Low local coherence, high macro coherence (IM)
  - High local coherence, low macro coherence (Lm)
  - High local coherence, high macro coherence (LM)
- Corpus: 21 texts, 100 dimensions
Foltz et al. (1998) experiment

- **Results:**
  - Coherence(lm) = 0.177
  - Coherence(IM) = 0.205
  - Coherence(Lm) = 0.210
  - Coherence(LM) = 0.242

![Graph showing comprehension vs. coherence with LSA r^2 = 0.853 and word overlap r^2 = 0.098]
Comparison of texts
Foltz et al. (1996)

• Compare texts and their summaries
• Britt et al. (1994) data
  • 24 participants read 21 texts
  • Participants write an essay
• Corpus: 21 texts + 8 encyclopedia texts + 2 book excerpts
• Assessing the essays:
  • *Score 1* = average similarity between each sentence and the closest text sentence
  • *Score 2* = same but compare with the main 10 text sentences
## Comparison of texts

**Foltz et al. (1996)**

<table>
<thead>
<tr>
<th>Judge1</th>
<th>Judge2</th>
<th>Judge3</th>
<th>Judge4</th>
<th>Score1</th>
<th>Score2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge1</td>
<td>.575**</td>
<td>.768**</td>
<td>.381</td>
<td>.418*</td>
<td>.589*</td>
</tr>
<tr>
<td>Judge2</td>
<td></td>
<td>.582**</td>
<td>.412*</td>
<td>.552**</td>
<td>.626**</td>
</tr>
<tr>
<td>Judge3</td>
<td></td>
<td></td>
<td>.367</td>
<td>.317</td>
<td>.384+</td>
</tr>
<tr>
<td>Judge4</td>
<td></td>
<td></td>
<td></td>
<td>.117</td>
<td>.240</td>
</tr>
</tbody>
</table>

** * p<.05  ** p<.01  + p<.06
Models based on LSA

- Semantic memory
- Vocabulary acquisition
- Text comprehension
A model of learning

- LSA learns the meaning of words from texts
- Does it mimic the human rate of learning?
- Landauer & Dumais (1997) experiment:
  - Most of the lexicon is acquired from text
  - Children data:
    - 3500 words read / day
    - 50 new words encountered / day
    - 10 words learned / day
    - Controlled experiments lead to 2.5 words learned / day
Direct and indirect effects

- How do children acquire the remaining 7.5 words/day?
- The poverty of stimulus problem
  - how do children acquire as much knowledge on the basis of little information?
- Hypothesis: a word is learned…
  - From texts containing it (direct effect)
  - From texts *not* containing it (indirect effect)
LSA simulation

- TOEFL test score is function of:
  - $T$: total number of paragraphs
  - $S$: number of paragraphs containing the stem word

\[
\text{score} = 0.128 \times (\log 0.076 \times T) \times (\log 31.910 \times S)
\]

- Correlation with observed score= .98
- After “reading” 25,000 paragraphs, what is the effect of the 25,001th?
  - When the stem word belongs to it?
  - When the stem word does not belong to it?
Example

- Suppose a word that appeared twice in the 25,000 paragraphs
  Score = 0.128 \( \log(0.076 \times 25000) \times \log(31.91 \times 2) \) = 0.75809

- Direct effect: the word belongs to the 25,001th paragraph
  Score = 0.128 \( \log(0.076 \times 25001) \times \log(31.91 \times 3) \) = 0.83202

- Indirect effect: the word does not belong to the 25,001th paragraph
  Score = 0.128 \( \log(0.076 \times 25001) \times \log(31.91 \times 2) \) = 0.75810
LSA simulation

- Effects are cumulated over all word-frequency band
  - Direct effect:
    - .0007 words gained per word encountered
    - \(.0007 \times 3500 = 2.5\) words acquired / day
  - Indirect effect:
    - .15 words gained per paragraph encountered
    - \(.15 \times 50 = 7.5\) words acquired / day
- The meaning of a word would be acquired
  - for 25% by reading text containing it
  - for 75% by reading texts not containing it
Direct and indirect effects
(Lemaire et al., 2006)

acheter (to buy)/magasin (store)
Cosine acheter/magasin

- Total gain: 0.58
- Gain due to the occurrence of acheter: -0.10
- Gain due to the occurrence of magasin: -0.19
- Gain due to the co-occurrence: 0.73
- Gain due to 2nd order co-occurrences: 0.11
- Gain due to 3rd and more co-occurrences: 0.03
Same measures on 28 words

- Total gain: 0.13
- Gain due to the occurrence of $W1$: -0.16
- Gain due to the occurrence of $W2$: -0.19
- Gain due to the co-occurrence: 0.34
- Gain due to 2$^{nd}$ order co-occurrences: 0.05
- Gain due to 3$^{rd}$ and more co-occurrences: 0.09
Models based on LSA

- Semantic memory
- Vocabulary acquisition
- Text comprehension
Text comprehension

- Connecting LSA as a model of semantic memory and the construction-integration model (Kintsch, 1998)
A Computational model of text comprehension (Lemaire et al, 2006)

Next sentence

- Select associates
- Retrieve previous elements

- SEMANTIC MEMORY (LSA)
- WORKING MEMORY
- EPISODIC MEMORY

Integration

Decay
Example (translated from French)

- A woodcutter was walking in the forest

- Select associates:
  - walk: stroll, meet, pick
  - woodcutter: ax, forest, cottage
  - forest: glade, oak, wood
Example (translated from French)

- A woodcutter was walking in the forest
A woodcutter was walking in the forest
A woodcutter was walking in the forest
A woodcutter was walking in the forest
The construction-integration model

- Before LSA, the construction step was performed by hand
- Link weights were set approximately
- All nodes being vectors, the weight of a link is the cosine of the corresponding nodes
- A question remains: what is the vector of \( P(A_1, \ldots A_n) \) ?
The predication algorithm
(Kintsch, 2001)

- The centroid rule:
  \[ P(A) \rightarrow \hat{P} + \hat{A} \]
- ...does not work well
  - The meaning of the predicate depends on the meaning of the arguments
  - The horse ran / The color ran
  - shark(lawyer) ≠ shark + lawyer

- A better rule:
  \[ P(A) \rightarrow \hat{P} + \hat{A} + \hat{M}_1 + ... + \hat{M}_n \]
  \( M_i \) are vectors that are close to both P and A
Example 1: construction step (Kintsch, 2001)

The horse ran ran(horse)

```
  ran
 /   \
|     |
|     |
|     | 21
|     |
  down
  /   \\
|     |
|     | 18
|     |
  hopped
  /   \\
|     |
|     | 12
|     |
  horse
```

(Kintsch, 2001)
Example 1: integration step

The horse ran \( \text{ran(horse)} \)

The horse ran \( \text{horse}+\text{ran}+\text{stopped} \)
Example 2: construction step

The color ran \textit{ran(color)}

- ran
- down
- hopped
- stopped
- color

P-values:
- .06
- .11
- .02
Example 2: integration step

The horse ran ran(horse)

The color ran color + ran + down
## Predication algorithm: test
(Kintsch, 2001)

<table>
<thead>
<tr>
<th>Landmarks</th>
<th>ran</th>
<th>horse ran</th>
<th>color ran</th>
</tr>
</thead>
<tbody>
<tr>
<td>gallop</td>
<td>.33</td>
<td>.75</td>
<td>.29</td>
</tr>
<tr>
<td>dissolve</td>
<td>.01</td>
<td>.01</td>
<td>.11</td>
</tr>
</tbody>
</table>
## Inferences (Kintsch, 2001)

<table>
<thead>
<tr>
<th>Predication</th>
<th>Centroid</th>
<th>Predication</th>
<th>Centroid</th>
</tr>
</thead>
<tbody>
<tr>
<td>The student washed the table</td>
<td>The student was clean</td>
<td>The table was clean</td>
<td></td>
</tr>
<tr>
<td><strong>Centroid</strong></td>
<td>.70</td>
<td><strong>Centroid</strong></td>
<td>.71</td>
</tr>
<tr>
<td><strong>Predication</strong></td>
<td>.62</td>
<td><strong>Predication</strong></td>
<td>.83</td>
</tr>
<tr>
<td>The student dropped the glass</td>
<td>The student was broken</td>
<td>The glass was broken</td>
<td></td>
</tr>
<tr>
<td><strong>Centroid</strong></td>
<td>.76</td>
<td><strong>Centroid</strong></td>
<td>.66</td>
</tr>
<tr>
<td><strong>Predication</strong></td>
<td>.87</td>
<td><strong>Predication</strong></td>
<td>.91</td>
</tr>
<tr>
<td>The hunter shot the elk</td>
<td>The hunter was dead</td>
<td>The elk was dead</td>
<td></td>
</tr>
<tr>
<td><strong>Centroid</strong></td>
<td>.66</td>
<td><strong>Centroid</strong></td>
<td>.54</td>
</tr>
<tr>
<td><strong>Predication</strong></td>
<td>.73</td>
<td><strong>Predication</strong></td>
<td>.70</td>
</tr>
</tbody>
</table>
Strengths and limits

- **LSA strengths**
  - fully automatic
  - vectorial representation allows:
    - easy representation of sequence of words
    - easy comparisons between words or texts
  - a cognitive model of learning and representation

- **LSA weaknesses**
  - no syntax processing
  - representations are not explicit
  - not incremental
  - paragraphs are bag of words
Cognitive models of semantic memory could mimic:
- the human semantic associations or
- the human *construction* of these associations over a long period of time

Semantic networks are good model of semantic associations BUT... they do not account for the construction of these associations
Features

- **Input** (corpus, word association norms)
- **Knowledge representation** (vector-based, network-based)
- **Addition of a new text** (incrementally or not)
- **Unit of context** (paragraph, sliding window)
- **Use of high-order co-occurrences** (yes/no)
- **Compositionality** (representing the meaning of a text from the meaning of its words)
Input: **corpus**
- Knowledge representation: **vector-based**
- Addition of a new text: **not incremental**
- Unit of context: **paragraph**
- Use of high-order co-occurrences: **yes**
- Compositionality: **easy**
Word co-occurrence matrix built from a N-word sliding window

Sample Global Co-occurrence Matrix for the Sentence
“the horse raced past the barn fell”

<table>
<thead>
<tr>
<th></th>
<th>barn</th>
<th>horse</th>
<th>past</th>
<th>raced</th>
<th>the</th>
</tr>
</thead>
<tbody>
<tr>
<td>barn</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>fell</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>horse</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>past</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>raced</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>the</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- vector = concatenation of row and column
- barn = <0,2,4,3,6,0,5,0,0,0>
- Measure of similarity = Euclidian distance
HAL (Burgess, 1998)

Table 2: Example Neighborhoods for *beatles* and *frightened*

<table>
<thead>
<tr>
<th></th>
<th>Neighbors</th>
</tr>
</thead>
</table>
| beatles | original  
          | hand            |
|        | song            |
|        | movie           |
|        | album           |
|        | songs           |
| frightened | scared       |
|            | upset           |
|            | shy             |
|            | embarrassed    |
|            | anxious         |
|            | worried         |
HAL
(Burgess, 1998)
HAL
(Burgess, 1998)

- **Input:** *corpus*
- Knowledge representation: *vector-based*
- Addition of a new text: *incremental*
- Unit of context: *sliding window*
- Use of high-order co-occurrences: *no*
- Compositionality: *easy*
Word association space
(Steyvers et al., in press)

- Based on association norms for 5,000 words
- Word/word matrix scaled to 200-500 dimensions
Figure 2. Correlations of different measures of semantic similarity for different dimensionalities. Data are taken from recognition memory, cued recall, free recall. See text for details.
Word association space
(Steyvers et al., in press)

- Input: association norms
- Knowledge representation: vector-based
- Addition of a new text: not incremental
- Unit of context: N/A
- Use of high-order co-occurrences: no
- Compositionality: easy
PMI-IR
(Turney, 2001)

- Based on Mutual Information (Church & Hanks, 89)
  - compares the probability of two words occurring together with the probability of the words occurring separately
  - \( \text{MI}(A,B) = \log_2 \left( \frac{p(A,B)}{p(A) \cdot p(B)} \right) \)

- TOEFL test: 1 problem word/4 choices
  - \( \text{score}_1(\text{choice}) = \frac{\text{hits}(\text{problem AND choice})}{\text{hits}(\text{choice})} \)
  - \( \text{score}_2(\text{choice}) = \frac{\text{hits}(\text{problem NEAR choice})}{\text{hits}(\text{choice})} \)
  - ...
**PMI-IR**
*(Turney, 2001)*

**Table 3. Results of the TOEFL experiments, including LSA results from [6].**

<table>
<thead>
<tr>
<th>Interpretation of $p(\text{problem} \mid \text{choice})$</th>
<th>Description of Interpretation</th>
<th>Number of Correct Test Answers</th>
<th>Percentage of Correct Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{score}_1$</td>
<td>co-occurrence using AND operator</td>
<td>50/80</td>
<td>62.5%</td>
</tr>
<tr>
<td>$\text{score}_2$</td>
<td>co-occurrence using NEAR</td>
<td>58/80</td>
<td>72.5%</td>
</tr>
<tr>
<td>$\text{score}_3$</td>
<td>co-occurrence using NEAR and NOT</td>
<td>59/80</td>
<td>73.75%</td>
</tr>
<tr>
<td>Latent Semantic Analysis</td>
<td></td>
<td>51.5/80</td>
<td>64.4%</td>
</tr>
<tr>
<td>Average Non-English US College Applicant</td>
<td></td>
<td>51.6/80</td>
<td>64.5%</td>
</tr>
</tbody>
</table>
PMI-IR
(Turney, 2001)

- **Input:** web
- Knowledge representation: N/A
- Addition of a new text: incremental
- Unit of context: web page
- Use of high-order co-occurrences: no
- Compositionality: hard
Network-based representation

- better for quickly retrieving neighbors
- better for representing asymmetrical relations

ICAN
(Lemaire & Denhière, 2004)
... if you have such a [device, connect the cable to the network connector] then switch ...  
Create or reinforce co-occurrence links
... if you have such a [device, connect the cable to the network connector] then switch ...

- Create or reinforce co-occurrence links
- Create or slightly reinforce 2nd-order co-occurrence links
... if you have such a [device, connect the cable to the network connector] then switch ...

- Create or reinforce co-occurrence links
- Create or slightly reinforce 2\textsuperscript{nd}-order co-occurrence links
- Slightly decrease other links
ICAN

- 3.2 million word corpus of texts for children
- Test on 1200 pairs of words (from association norms)
- Average weights:
  - stem/1\textsuperscript{st} associate: .415
  - stem/2\textsuperscript{nd} associate: .269
  - stem/3\textsuperscript{rd} associate: .236
  - stem/last associates: .098
- Correlation with human data: .50
ICAN

- **Input:** corpus
- **Knowledge representation:** network-based
- **Addition of a new text:** incremental
- **Unit of context:** sliding window
- **Use of high-order co-occurrences:** yes
- **Compositionality:** hard