

- 1 Applied Psycholinguistics
- 2 Applied Psycholinguistics
 - Some Topics in Applied Psycholinguistics
 - Resources
 - LiveInk
 - Latent Semantic Analysis
- 3 Some Topics in Applied Psycholinguistics
 - 1
 - Reading Instruction
 - Spelling
 - Text Design
 - Diagnosis and Treatment
 - Measures of Readability
 - Writing Assessment
 - 2
 - Text Formatting
 - Machine Translation
 - Automatic Speech Recognition
 - Information Retrieval
 - Second Language Learning
 - Natural Language Interfaces
- 4 Resources
 - Applied Psycholinguistics
 - Annual Review of Applied Linguistics
 - Computational Linguistics
- 5 LiveInk
 - What is LiveInk?
 - Prior Research
 - Linzie & Fletcher
 - Fletcher, Domingo & Sung
- 6 What is LiveInk?
 - Can you read this?
- 7 What is LiveInk?
 - What about this?
- 8 What is LiveInk?
 - Is this better?
- 9 What is LiveInk
 - Why is the second example easier to read?
 - Capitalization
 - Punctuation

- Spacing
- Why might the third be even easier?
 - Line breaks at syntactic boundaries.
 - Indentation shows relationships among constituents.
 - Verbs are highlighted.
 - Sentences are presented one at a time.

10 **What is LiveInk?**

- Walker Reading Technologies, Inc.
- Goal is to use technology to improve reading.
- Algorithms are proprietary and technology is patented.
- Website: <http://www.liveink.com/>

11 **Prior Research**

- Graf & Torrey (1966)
- Reading times are faster when line breaks correspond to major phrase structure boundaries.
- Breaks at Major Boundaries
 - "During World War II, even fantastic schemes received consideration if they gave promise of shortening the conflict."
- Breaks Not at Major Boundaries
 - "During World War II, even fantastic schemes received consideration if they gave promise of shortening the conflict."

12 **Prior Research**

- Educational Research on "Linguistic Cueing"
 - Use "spatial cues" to signal grammatical constituents or thought units.
 - Faster RT in an information finding task (Frase & Schwartz, 1979)
 - Faster reading times (North and Jenkins, 1951)
 - Improved comprehension (North and Jenkins, 1951)
 - Results depend on text difficulty and reader ability.

13 **Prior Research**

- In spoken language, prosodic cues (intonation, stress) are used to signal grammatical boundaries and relationships between constituents.
- When these cues are removed, comprehension is impaired (Cutler, Dahan & Donselaar, 1997).
- These cues are lost when language is written down.

14 **Linzie & Fletcher**

- Compared LiveInk to normal blocked presentation.
- Participants were college students.
- Narratives (Aesop's fables) and expository texts (from the Federal Register) were tested.
- Dependent variables were
 - Reading Rate
 - Comprehension Questions
 - Preference Ratings

15 **Linzie & Fletcher**

- LiveInk slows readers down.
- LiveInk has little impact on the comprehension of narratives.

- LiveInk improves comprehension of expository texts.
- Readers prefer LiveInk.
- Problems?

16 Fletcher, Domingo & Sung

- Goals:
 - Is the slowing down we see with LiveInk caused by all the line breaks?
 - Do college students benefit when reading more “normal” texts?
 - Do the benefits we see with LiveInk depend on where we place the line breaks, indents and highlights?
 - Do some people benefit more from LiveInk than others?

17 Fletcher, Domingo & Sung

- Participants and Materials
 - Participants are U of M psychology students. 64 native speakers and 24 non-native speakers.
 - Materials are 16 texts taken from the GRE reading comprehension sub-test.
 - One important and one unimportant sentence identified within each.

18 Fletcher, Domingo & Sung

- Independent Variables
 - Cascaded versus Blocked Presentation Format
 - controlling for line breaks
 - LiveInk versus VileInk
 - Important versus Unimportant Sentences
 - Native Versus Non-Native Speakers

19 Fletcher, Domingo & Sung

- Procedure
 - Read 8 texts (plus warmup) in one format, rest, then read 8 texts (plus warmup) in the other format.
 - After reading each text, reread and fill in the blanks.
- Dependent Variables
 - Reading Times
 - Cued Recall

20 Fletcher, Domingo & Sung

- Summary of Results
 - No difference in reading time between LiveInk and VileInk.
 - No difference in cued recall for native speakers.
 - Significant difference in free recall for non-native speakers.
 - Greatest difference for important statements.

21 Latent Semantic Analysis

- What is LSA?
- LSA and Multidimensional Scaling
- Applications of LSA

22 What is LSA?

- “**Latent Semantic Analysis (LSA)** is a theory and method for extracting and representing the contextual-usage meaning of words by statistical computations applied to a large corpus of text.”
- Starts with a **large** corpus of text.
 - English and American literature of the 18th and 19th centuries from Project Gutenberg. (57,092,140 words!)
 - Encyclopedia (60,768 unique terms)
 - All the text from three college level psychology texts.
 - Closed captioning data from Disney Channel programming.

23 What is LSA?

- The similarity of all possible pairs of words is calculated by computing the similarity of the contexts in which they occur.
 - “Couch” and “sofa” are similar in meaning because they occur in similar contexts.
- A mathematical procedure similar to factor analysis is used to discover the underlying semantic space.
- A typical semantic space has 50 – 1,000 dimensions!

24 What is LSA?

- The meaning of any word can be represented as its location in this semantic space.
- The meaning of a sentence or text is represented as the central tendency of the words that make it up.
 - “The dog bit the man” = “The man bit the dog.”

25 What is LSA?

- Major References
 - Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.
 - Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- <http://lsa.colorado.edu/>
 - Demos
 - Applications
 - Publications

26 LSA and Multidimensional Scaling

- Where do the semantic distances come from?
 - Similarity Ratings
 - Automatic Analysis of Contexts
- Dimensionality of the Solution

27 Applications of LSA

- Scoring Essay Exams
- Educational Text Selection
- Information Retrieval

28 Scoring Essay Exams

- LSA to analyze the quality of 1,205 GMAT essays by determining how close they were in semantic space to previously graded essays.
- Results were compared to those assigned by two experienced ETS graders.
 - Grader 1 – Grader 2: $r = .71$

– LSA – Grader: $r = .70$

29 **Educational Text Selection**

- Learning from text depends on the match between the readers background knowledge and the difficulty of the text.
 - Zone of Learnability
- Participants read one of 4 texts about the human heart.
 - Children’s Book on the Human Body
 - General Introduction to the Circulatory System
 - Undergraduate Anatomy Text
 - Medical School Text

30 **Educational Text Selection**

- Pre- and post-tests were used to assess how much participants learned.
 - Short Answer Questionnaire
 - Essay on the Functioning of the Heart
- LSA was used to compare the pre-test essays to the chapter from the undergraduate anatomy essay.
- How much participants learned from each text was systematically related to how much they knew pre-experimentally as measured by LSA.

31 **Information Retrieval**

- Search engines such as Google use keywords or literal phrases to retrieve information.
- LSA can be used to measure how similar candidate texts (e.g. web pages) are to a user’s query or to an example text.
- Research has shown that LSA does improve search results.

32 **The End!**