Brückenkurs 2006

Introduction to Psycholinguistics

Pia Knoeferle & Matthew W. Crocker

Department of Computational Linguistics, Saarland University

October 13, 2006

Overview

- What is psycholinguistics?
- A brief history of psycholinguistics
- Core issues: Modularity
- Language processing
  - Incrementality
  - Ambiguity in language
  - Language-vision interaction
- Experimental design and analysis

Psycholinguistics

- Some definitions
  - "Study of the psychological / neurobiological factors that enable us to acquire, use, and understand language"
    - Language acquisition
    - Language production / comprehension
  - "To understand and model the processes that underlie the human capacity to process language"
    - How does the human language processor work?
    - How is it realized in the brain?
    - How is linguistic knowledge represented in the brain?
    - Where does our capacity for language emerge from?

A brief history

- A brief history (see Saffran, 2003)
  - In Greece & Rome, capacity for language was ascribed to the tongue
- *Phrenology* (Gall, ca. 1800):
  - Bumps on the skull taken to reflect areas of enlargement in the brain
  - Located language at the protrusion of the eye socket below the eye
- Empiricism of the 18th/19th century
  - Study of *aphasia* (partial or total loss of the ability to articulate ideas or comprehend spoken or written language) following brain damage

Organization of language in the brain

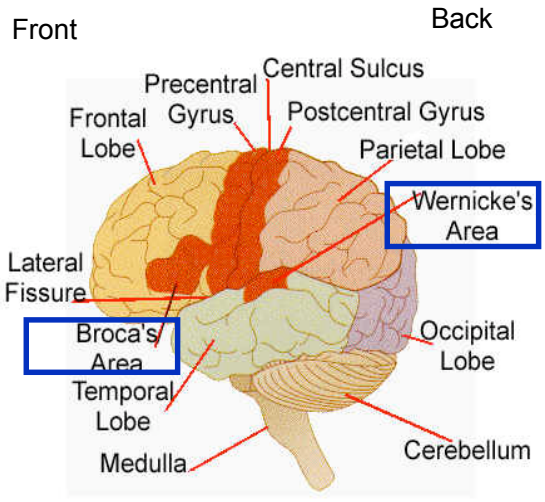
- Paul Broca (physician, ca. 1860)
  - Localization of functions in the cerebral cortex
    - M. Leborgne, suffered a stroke, only uttered one monosyllable
    - Autopsy: area of damage in left inferior frontal lobe (Broca's area)
    - Broca's area lies anterior to area of the motor cortex that transmits commands to face muscles, tongue, and larynx
- Broca's aphasia
  - Sparse speech, nonfluent
  - Intonation and stress patterns are deficient
  - Lack of grammatical structure
    - Sentence construction is poor
    - Disjointed words
    - Omitting function words and inflections
      - E.g., *Son ... University ... Smart ... Boy ... Good ... Good ...*

Organization of language in the brain

- Carl Wernicke (German physician)
  - Damage to an area of the left superior temporal lobe (part of auditory association cortex, and next to primary auditory cortex)
  - Result: Loss of language comprehension
    - Wernicke's aphasia
      - Deficits in comprehension and repetition
      - Speak fluently but content is often incorrect
      - Difficulty in word retrieval, generation of neologisms
  - Superior temporal area: where auditory word images are stored
  - Proposed a connection between Wernicke and Broca

  - Example (Saffran, 2003, Appendix 1)
    - [...] I can't recall the word that I can't thay. With the French, you know, uh, what is the name of the word, God, public serpinz they talk about, uh, but I have had that, it was ryediss, just before the storage you know, seven weeks, I had personal friends that, that, I would cook an' food the food and serve fer four or six mean for an evening.

Organization of language in the brain



http://www.tbts.org/itemDetail.asp?categoryID=295&itemID=16377

- **Frontal lobe**
  - Separated from parietal lobe by central sulcus/rolandic fissure
  - **Contains Broca's Area**
    - Damage can result in impaired language production
- **Temporal lobe**
  - Separated of frontal and parietal lobes by the lateral/sylvian fissure
  - **Wernicke's area**
    - Damage can result in problems processing auditory language
- Occipital lobe
  - Visual processing
  - Damage can impair processing of written language

Human language processing: Function

- What does it do?
  - Comprehension: Maps from "sound to meaning"
    - speech/orthography to words
    - words to structures
    - structure to meanings
  - Production: Maps from "message to speech"
    - Meaning to grammatical encoding
    - Phonological encoding
    - Articulation

Competence versus Performance

- Competence: Knowledge of Language
  - Linguistic theories at all levels
    - Phonetics/phonology, morphology, syntax, semantics ...
  - Rules and representations

- Performance: How Language is Processed
  - Use of Knowledge of Language
    - Processes for comprehension and production
  - Architectures and Mechanisms

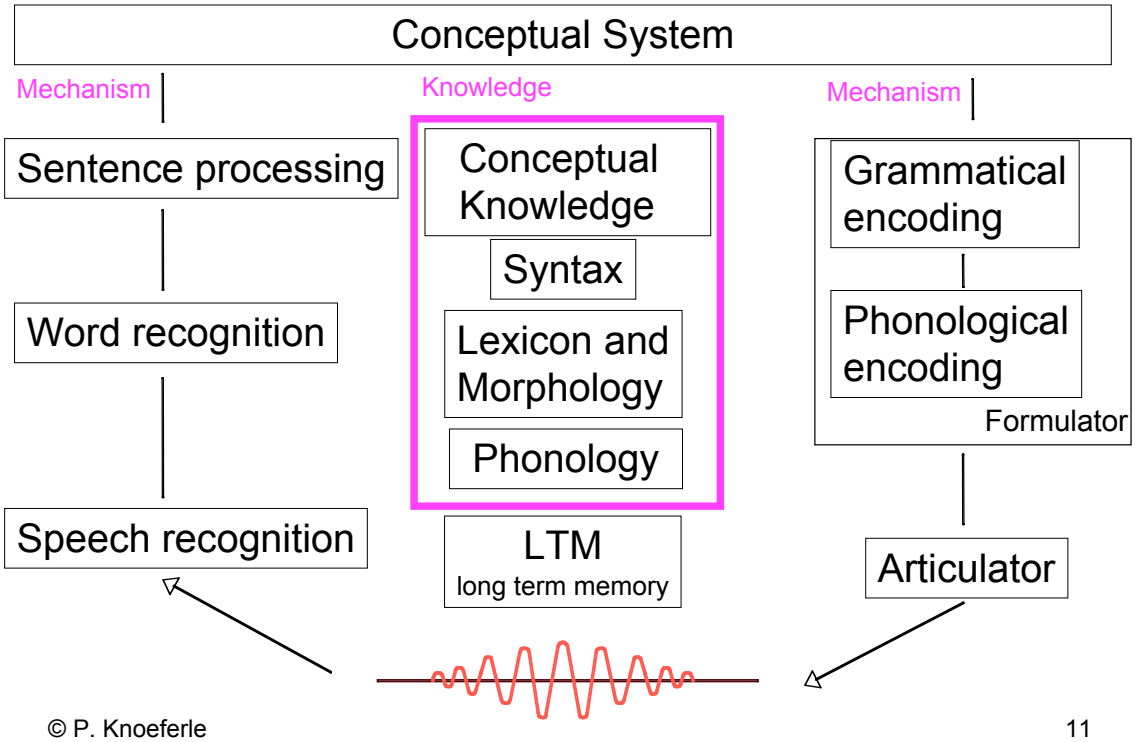
Why Distinguish Competence & Performance?

- Sometimes what we do differs from what we know
- Production: we say things we know are wrong
  - Spoonerisms: “Mental lexicon” spoken as “Lentil Mexican”
  - Agreement: “The friend of the two girls are laughing”
- Comprehension: we can’t understand things we know are ok
  - Centre embedding:
    - “The mouse that the
  - Garden paths:
    - “The horse raced pas

More Spoonerisms

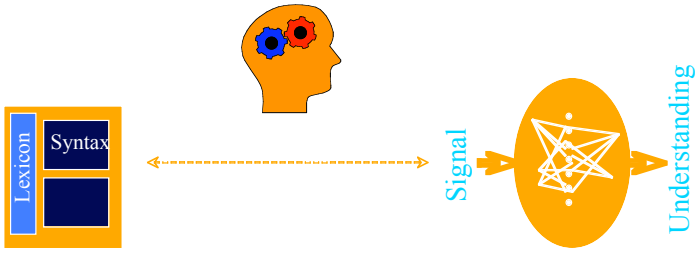
- A lack of pies (A pack of lies)
- Wave the sails (Save the whales)
- Plaster man (Master plan)
- Bottle in front of me (Frontal Lobotomy)
- Rental Deceptionist (Dental Receptionist)
- Flock of bats (Block of flats)
- Chewing the doors (Doing the chores)

Speech Processing Model (Dijkstra & Kempen, 1993)



The Modularity Issue

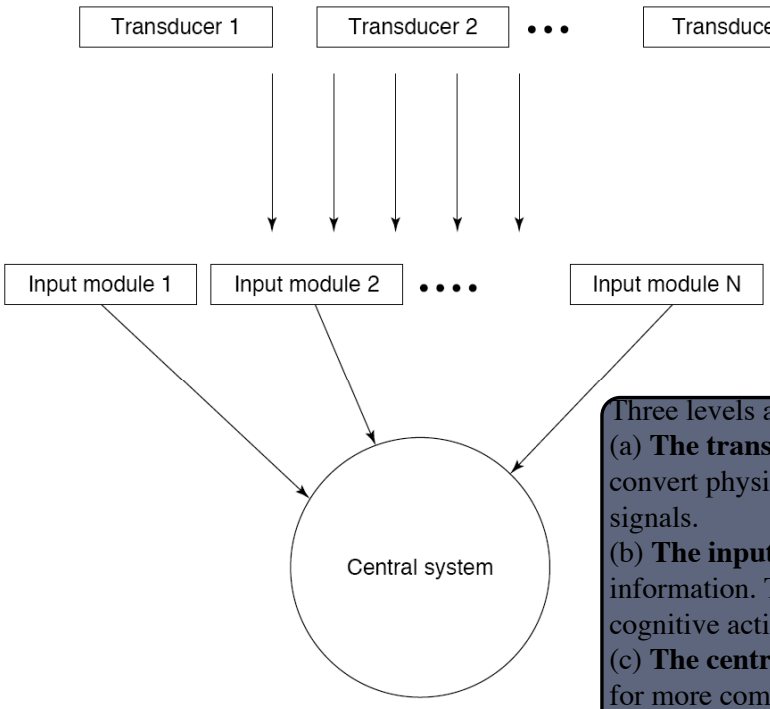
- Is language distinct from other other cognitive processes?
  - ➪ e.g. vision, smell, reasoning ...
- Do distinct modules exist within the language processor?
  - ➪ e.g. word segmentation, lexical access, syntax ...
- What is a module anyway!?



Architectures and Mechanisms

- What does “distinct” mean:
  - Representational autonomy
    - E.g. phonological representations differ from syntactic representations
    - Possibly interactive processes
  - Procedural autonomy
    - E.g. lexical access versus syntax
    - Possibly shared representations
- How are any such “distinct subsystems” for language processing organised? How do they interact?
  - Does organisation affect possible mechanisms?
  - Theoretical, computational and empirical arguments for and against ‘modularity’?

Fodor’s Modularity



Modules are, e.g.:
- domain specific
- innately specified
- informationally encapsulated
- fast
- autonomous

Three levels are distinguished:
(a) **The transducers**, whose function is to convert physical stimulation into neural signals.
(b) **The input systems**, interpret transduced information. They are responsible for basic cognitive activities and are modular.
(c) **The central system**, is responsible for more complex cognitive activities such as analogical reasoning, and is not modular.

Reference
a Fodor, J.A. (1983) *The Modularity of Mind*, MIT Press/Bradford Books

Summary

- Definition of ‘psycholinguistics’
- Brief history of early days in psycholinguistics
  - ➪ Language organization in the brain
- Function of language processing
  - ➪ Comprehension and production
- Competence vs. performance and reasons for differentiating between them
- Modularity issue
  - ➪ Encapsulated modules might account for the rapidity of language processing

Human Language Processing

- We understand language incrementally, word-by-word
  - ➪ How do people construct interpretations?
- We must resolve local and global ambiguity
  - ➪ How do people decide upon a particular interpretation?
- Decisions are sometimes wrong!
  - ➪ What information is used to identify we made a mistake?
  - ➪ How do we search for an alternative?

Real-time language processing

- Speech streams include no discrete boundaries to indicate where one word ends and another begins.
- We understand stammering non-fluent politicians and non-native speakers. Incomplete sentences are no problem for us.
- We deal with ambiguity all the time without breaking down. Computer parsers often maintain thousands of possible interpretations.
- We have a vocabulary of about 60,000 words. We access somewhere between 2-4 words/second (error rates around 2/1000 words)
- We understand speech even faster than we can produce it. We are so fast, we can even finish each others sentences.

Investigating real-time language processing

- How do we know people understand language incrementally?
  - ➪ Speech shadowing task: the participant repeats back speech as he hears it
    - Close shadowers (~10% of people) can repeat what they hear at a delay of only 250 ms (normal ~500 ms)
    - 250 ms = 1 syllable, i.e. close shadowers are processing the incoming material at the level of individual syllables

Marslen-Wilson, *Nature*, 1973

Investigating real-time language processing

- Ambiguity has often been exploited to investigate claims about language comprehension
- Types of ambiguity
  - Lexical ambiguity
  - Structural ambiguity
    - Word order and grammatical function
    - Thematic role ambiguity
    - Attachment ambiguity
  - Referential ambiguity
  - Local/global ambiguity
  - Multiple ambiguity

Ambiguity

- Lexical (at word level)
  - Word category
    - E.g., verb/noun: *walk, praise, call, phone*
  - Meaning
    - *Homographs*: same written form but different meaning
      - *E.g., Lauf* (Klavier, Joggen), *Geschirr* (Pferd, Porzellan)
      - *Peter remembered the ball which he had lost yesterday.*
      - *Peter remembered the ball which he had gone to last week.*
    - *Homophones:* same pronunciation, different in meaning
      - *Tale/tail; bear/bare*;

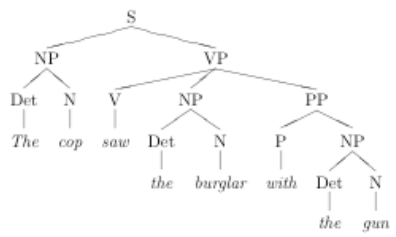
Ambiguity

- Structural ambiguity
  - Word order: subject-object (SO) vs. object subject (OS)
    - *Die Mutter grüßt die Tante.*
      - Word order variation: Subject-object and object-subject are grammatical
      - Case ambiguity: for feminine nouns NOM and ACC case are identical
  - Thematic role ambiguity
    - *The princess called by the king was talking to her nurse.*
    - The first noun phrase is both
      - Agent (main clause *The princess was talking to her* nurse)
      - Patient (reduced relative clause *called by the king*)

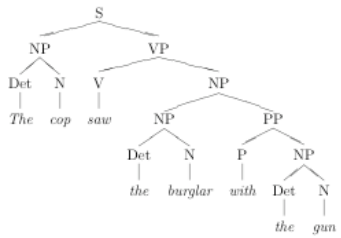
Ambiguity

- Structural ambiguity
  - Attachment ambiguity
    - PP-attachment

(1) VP-attachement

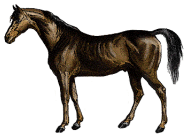


(2) NP-attachment

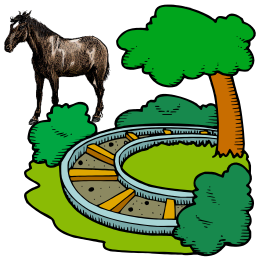


Ambiguity

- Referential ambiguity
  - *The horse was beautiful.*



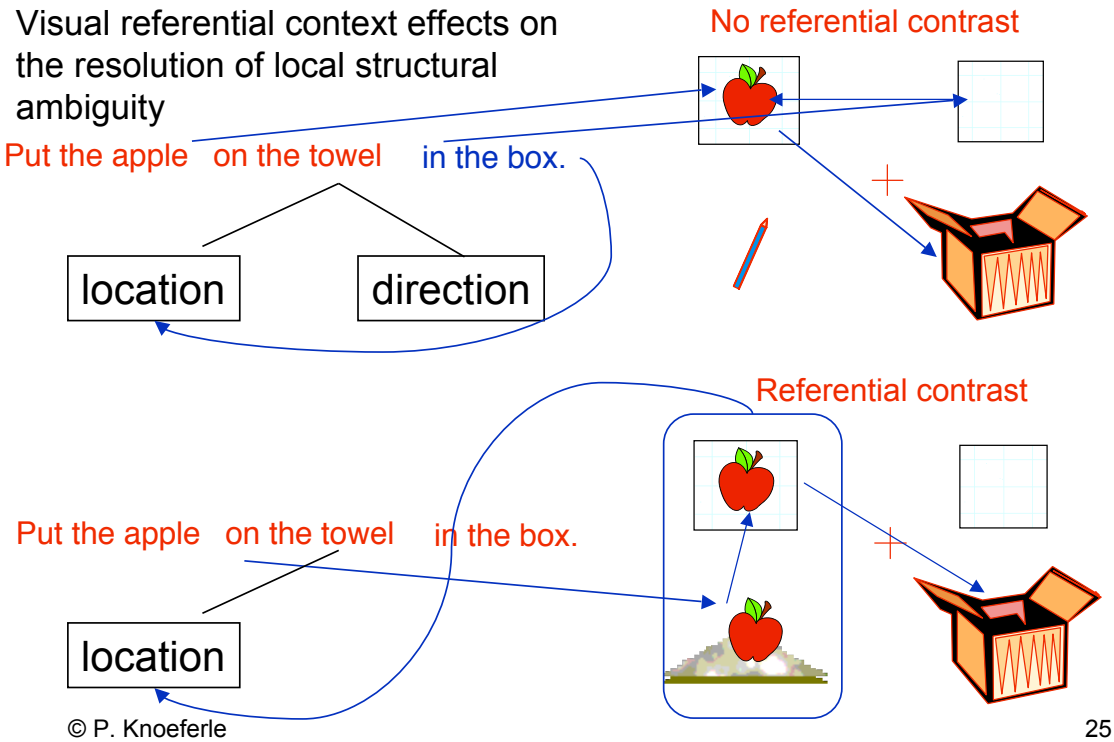
  - *The horse next to the tree was beautiful.*



Ambiguity

- *Local* versus *global* ambiguity
  - Disambiguating information
    - Information that resolves a local ambiguity
      - *Die Frau grüßt den Mann* (local)
      - *Die Frau grüßt die Tante* (global)
- Multiple ambiguity (more than two structural analyses)
  - *The girl hit the boy with the book on the hill.*
    - The book was an instrument for hitting the boy or the boy had the book
    - The hill was either the location of the book or of the hitting

Against modularity: Tanenhaus et al., 1995, *Science*



Summary

- People process language incrementally
- They understand language efficiently and rapidly
  - And this despite language being rich in ambiguity
    - Lexical, structural, referential
    - Sometimes the ambiguity is local, sometimes global
  - E.g., rapid use of visual referential context to resolve local structural ambiguity (evidence against strict modularity)

Experimental design

- We have just heard about the Tanenhaus study ...
- The study tested whether a visual referential context can influence the initial structuring of an utterance
- So what if we wanted to test a theory ourselves we could conduct an experiment
  - *Experiment* (contrasted with *non-experiment:* introspection, data collection)
    - Some definitions of what an experiment is
      - Systematic observations of a specific behaviour under controlled circumstances
      - Set of actions and observations, performed to verify or falsify a hypothesis or research a causal relationship between phenomena
      - The act of conducting a controlled test or investigation

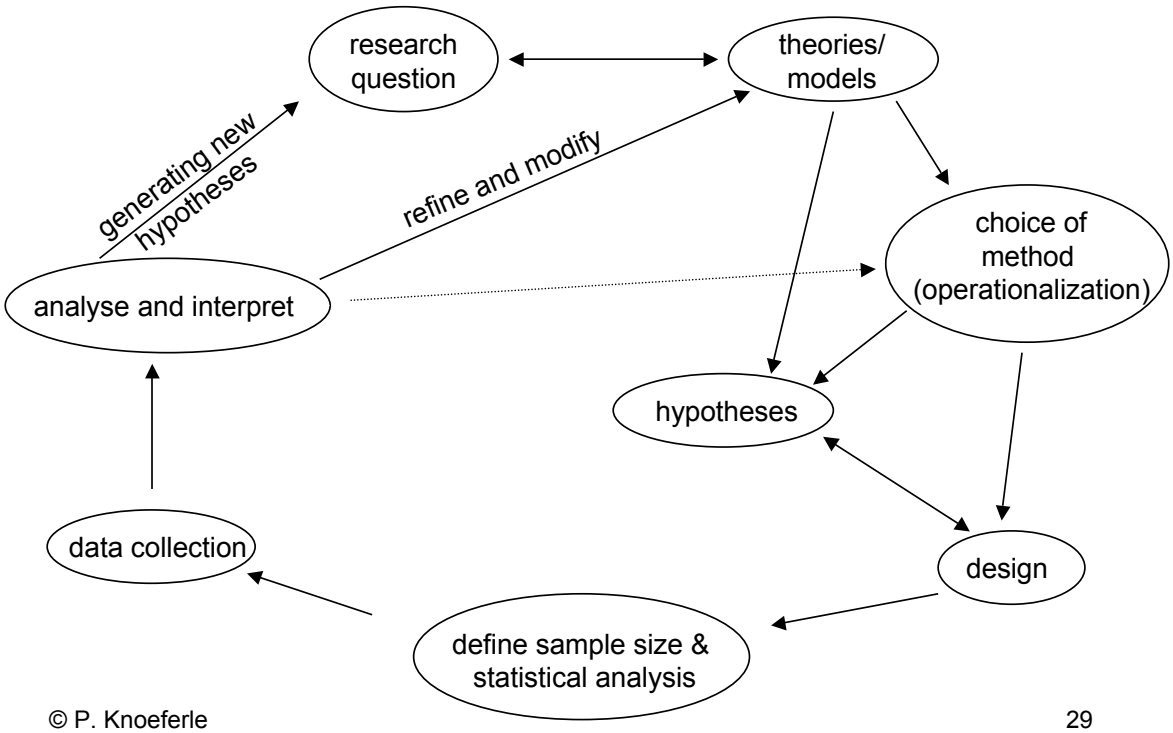
Experimental design

- Experimental designs
  - Example
    - Effects of preparation on using hedges versus assertions

	Hedging expressions	Assertive statements
Prepared	(1)	(1)
Unprepared	(2)	(2)

  - All variables can be freely manipulated
  - Participants can randomly be assigned to the experimental conditions
  - Avoids systematic effects of other (correlated) variable

Experimental design



Experimental design

- We could, for instance, use what we have heard so far about ambiguities to test some claims about incremental syntactic structure building

- Research question
  - How does the comprehension system build a syntactic analysis of a sentence?
  - ➪ Theory1 (simplest-first)
    - We follow strategies (e.g., choose the simplest analysis first)
  - ➪ Theory2
    - We do not chose the simplest analysis first

Experimental design

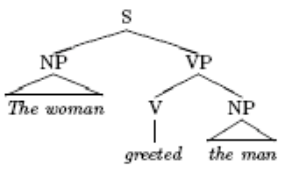
- Operationalize
  - If simplest-first theory is true, people should experience processing difficulties when simplest analysis is disconfirmed
- Hypotheses
  - H1 (*Experimental hypothesis*)
    - Processing difficulty when simplest analysis disconfirmed
  - H0 (*Null hypothesis*)
    - No processing difficulties when simplest analysis disconfirmed
- Design
  - Independent variable(s)
    - What we manipulate
  - Dependent variable(s)
    - What we measure

Experimental design

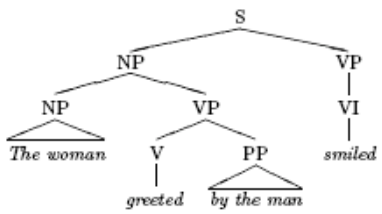
- Design
  - 1 *Factor* design:
  - Factor 1: Sentence type (2 *levels*: simple (a) vs. complex (b) analysis)

(a) The woman greeted the man and smiled.
(b) The woman greeted by the man smiled.

(a)



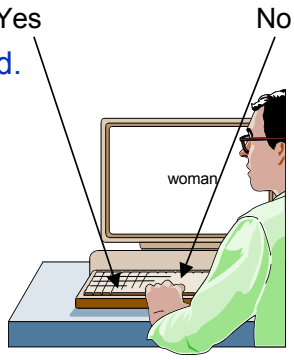
(b)



  - Design/materials issues
    - Match the length and frequency of the words in your 2 conditions
    - Confounds (e.g., differences in plausibility)? -> Counterbalancing

Methods

- Lexical decision, naming, window-methods in self-paced reading (SPR), eye tracking, neuropsychological methods
  - ➪ Lexical decision
    - (a) The woman greeted the man and smiled.
    - (b) The woman greeted by the man smiled.
  - ➪ Pros/cons of using this method?



Methods

- Self-paced reading:
  - ➪ Present a sentence one word/sentence region at a time on the screen
  - ➪ After reading that word, people click a button to see the next word(s)
  - ➪ Pros
    - Fairly incremental - per-region reading times
  - ➪ Cons
    - Only total reading times in a region (i.e., no fine-grained measures)
    - You only see one region at a time: artificial presentation that does not allow you to re-read earlier text
- Eye tracking
  - ➪ Pros
    - Incremental, per-region reading times
    - Fine-grained distinctions: first pass (time spent in a region when first entering it, and before leaving it), and total times
    - The entire sentence is presented, allowing people to re-read text
  - ➪ Cons
    - No direct evidence of neural activation, for instance

Statistical tests

- Once we have designed the experiment, created the materials, and run the study, we want to analyse the recorded data: statistics
- In our study we had 2 conditions
- We want to ascertain whether any differences in reading times between the MC and the RR condition are *systematic* or *unsystematic*
  - Systematic variation: result of experimental manipulation
    - E.g., MC vs RR sentence condition
  - Unsystematic variation: due to random factors: e.g., age, gender
- Test statistics (e.g., F-statistic, Analysis of Variance: ANOVA)
  - Discover how much variation there is in performance
  - How much of this variation is systematic versus unsystematic
  - Is there more variation than without the experimental manipulation?

ANOVA

- An ANOVA tells you whether
  - Differences between conditions are due to your manipulation
  - Due to unsystematic variation
- The ANOVA can help us analyse differences between means in more complicated designs (e.g., with 2 or more factors)
  - The result of an ANOVA analysis is a F-value
    - Ratio of the variance due to your experimental manipulation over unsystematic variation
    - A high F-value indicates a lot of the variation results from your manipulation

$$F = \frac{\text{systematic variation}}{\text{unsystematic variation}}$$

Summary

- Psycholinguistics deals with language processing
- A brief history of psycholiguistics
  - ➪ Broca's and Wernicke's aphasia and language areas in the brain
- Competence vs. performance (comprehension & production)
- Core issues: Modularity
  - ➪ Fodor (1983)
- Incrementality of comprehension
- Ambiguity in language and its uses in psycholinguistics
  - ➪ Lexical, syntactic, referential ambiguity; global vs. local;
- Experimental design and analysis