

Computational Psycholinguistics

Empirisch-methodische Grundlagen

Typische Fragestellungen (bei weitem keine vollständige Liste!...)

⇒ Sprachverstehen

- ⇒ Wie werden grammatikalische Mehrdeutigkeiten aufgelöst? z.B.:
Peter beobachtete den Spion mit dem Fernglas (globale Ambiguität)
Peter traf die Professorin, die die Studentinnen suchten. (lokale Ambiguität)
- ⇒ Parallele vs. serielle Verarbeitung.
- ⇒ Wie interagieren verschiedene Informationsquellen (Lexikon, Syntax, Semantik, Pragmatik, Weltwissen etc.) miteinander?
- ⇒ Rolle von Spracherfahrung und kognitiven Restriktionen (z.B. Arbeitsgedächtnis).

⇒ Sprachgenerierung (Produktion)

- ⇒ Wie lassen sich systematische (dennoch seltene) 'Fehler' erklären (z.B. *the friend of the busy students were ill*)
- ⇒ Welche Faktoren bestimmen die Auswahl einer Satzstruktur (z.B. *I gave him the book* vs. *I gave the book to him*)

etc.

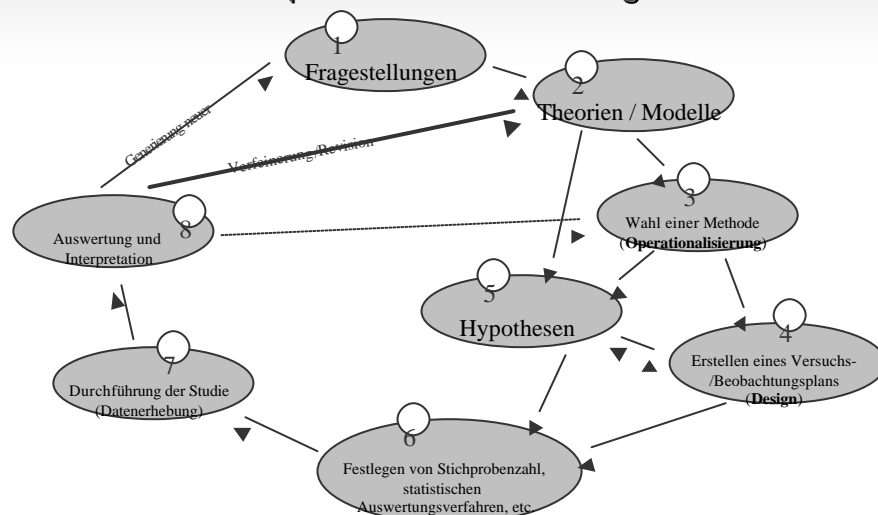
Warum messen/experimentieren?

- Linguistische Intuition (Introspektion) ist notorisch **subjektiv**
- Abstufungen (z.B. hinsichtlich der Akzeptabilität von Sätzen/Aussagen) sind intuitiv schwer meßbar
- Prozeßverläufe während der Satzverarbeitung sind intuitiv kaum zugänglich (vgl. **on-line/off-line** Kontrast bei *Manfred beobachtete den Spion mit dem Schnurrbart mehrere Minuten lang*).

daher

- Vergleich der Verarbeitung verschiedener (theoretisch relevanter) Satzkonstruktionen im psycholinguistischen Experiment
- Stichproben bestehend aus mehreren ('naiven') Versuchspersonen
- Wahl bestimmter experimenteller Techniken, die Aussagen über Akzeptabilitätsabstufungen bzw. Prozeßverläufe zulassen
- Ausschließen bzw. Kontrolle möglicher 'Störfaktoren' und Alternativerklärungen (Maximierung der *internen Validität*) – durch entsprechende Versuchsplanung und Pre-Tests.

„Kreislauf“ empirischer Forschung



Beispiel

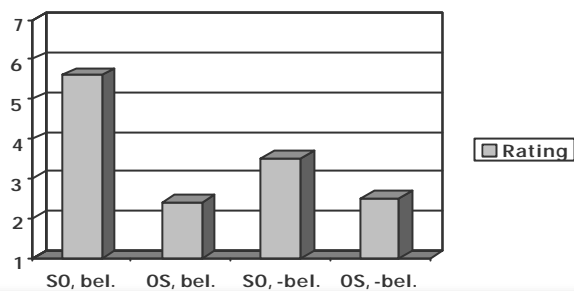
- **Fragestellung:** Wie werden morphologisch ambige NPen am Satzanfang (e.g. Daß die Frau ...) analysiert?
- **Theorie1:** Gar nicht (Unterspezifikation)
Theorie2: Doch! Und zwar als Subjekt (weniger komplex, häufiger)
Theorie3: Als Subjekt nur, wenn *belebt* – ansonsten als Objekt.
- **Operationalisierung:**
Wenn spätere Information (z.B. Kasus einer zweiten NP) die syntaktische Rolle von NP1 desambiguiert, sollten sich u.U. **Verarbeitungsschwierigkeiten** ergeben (*Garden path* Effekt)
HT1: nie!
HT2: Ja, wenn NP2 als eindeutiges Subjekt fungiert
HT3: Ja, wenn NP1 belebt und NP2=eindeutiges Subjekt, bzw. wenn NP1 unbelebt und NP2=eindeutiges Objekt.

Beispiel (continued)

- **Methode:** Akzeptabilitätsurteile (je schwieriger zu Verarbeiten, desto geringer die Akzeptabilität), Skala von 1=gar nicht akzeptabel – 7=völlig akzeptabel)
- **Versuchsplan (Design)**
 - (1) Daß die Frau den Mann stört, ist unschön. (SO, belebt)
 - (2) Daß die Frau der Mann stört, ist unschön. (OS, belebt)
 - (3) Daß die Ruhe den Mann stört, ist unschön. (SO, -belebt)
 - (4) Daß die Ruhe der Mann stört, ist unschön. (OS, -belebt)
- **Hypothesen:**
 - HT1:** (1) = (2) = (3) = (4) (kein Effekt)
 - HT2:** (1) > (2), (3) > (4) („Haupteffekt“ Abfolge)
 - HT3:** (1) > (2), (3) < (4) („Wechselwirkung“ Abfolge x Belebtheit)

Beispiel (continued)

- **Stichprobe:** 40 Versuchspersonen, 8 Sätze je Bedingung pro Person (= 320 Datenpunkte pro Bedingung)
- Statistisches Verfahren:** Repeated Measures ANOVA
- Signifikanzniveau:** 5%
- **Durchführung** (z.B. Fragebogen im Internet)
- **Auswertung:**
Deskriptive Analyse



Beispiel (still continued)

- **Inferenzstatistische Analyse:**
 - (1) > (2): signifikant ($p < 5\%$)
 - (3) > (4): signifikant
 - und unerwarteterweise: (1) – (2) > (3) – (4) signifikant (Wechselwirkung)
 - **Interpretation:**

Theorie1 und **Theorie3** können (in ihrer strengen Form) verworfen werden, **Theorie2** paßt am besten zu den Daten.

Jedoch: Einfluß von Belebtheit der ersten NP, den **Theorie2** ebenfalls nicht erklären kann.
- => **Theorie-Erweiterung (?)**, Experimente mit besser kontrolliertem Material (?), Experimente mit anderer Methode (?)...

Experimentelle Techniken

- ⇒ **Nicht-experimentelle Methode** (keine *Manipulation*, sondern lediglich *Beobachtung* der Zusammenhänge zwischen verschiedenen Einflußgrößen):
z.B. **Korpusanalyse**
- ⇒ **Off-Line Methoden**, z.B.
 - ⇒ Akzeptabilitätsurteile (siehe Beispiel)
 - ⇒ Sentence Completion (Satzvervollständigung), z.B. *Daß die Frau ____* . vs. *Daß die Ruhe ____* (kombiniert Verstehens- und Generierungsaspekte)
(Off-line Methoden erlauben lediglich, Aussagen über das **Ergebnis** von Verstehens/Generierungsprozessen zu Treffen)
- ⇒ **On-Line Methoden**, z.B.
 - ⇒ Self Paced Reading (Reaktionszeiten)
 - ⇒ Eyetracking (Blickbewegungsmessung)
(On-line Methoden erlauben Aussagen über den **Verarbeitungsverlauf** => **Inkrementalität** der Verarbeitung!)

Self Paced Reading

- ⇒ Versuchspersonen bekommen das Satzmaterial in Segmenten (Wörter oder Wortgruppen) dargeboten.
- ⇒ Per Tastendruck wird ein Segment aufgerufen, gleichzeitig wird ein zuvor dargebotener Abschnitt maskiert (nicht-kumulative Darbietung)
- ⇒ Die Lesegeschwindigkeit bestimmt die VP selbst (=> *self paced*)
- ⇒ Gemessen wird die Zeit zwischen den Tastenreaktionen, als Maß für die Verarbeitungsschwierigkeit am jeweiligen Verarbeitungspunkt.

Beispiel

--- die Frau --- ---- - - - - -

- ⇒ **Vorteile:** Billige, leicht zu realisierende On-line Methode
- ⇒ **Nachteile:** Eher „unnatürliche“ Darbietung; Einflüsse der Segmentierung möglich.

Eyetracking

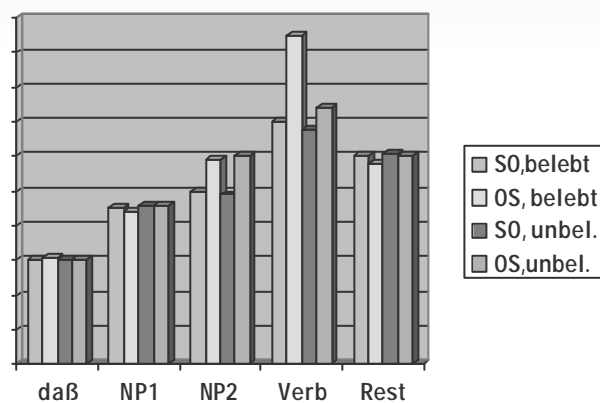
- Versuchspersonen lesen Sätze auf „natürliche Weise“ (keine Segmentierung des Materials), während Blickbewegungen aufgezeichnet werden.
- Messung verschiedener Variablen: Fixationszeiten auf Wörtern im ‚ersten Durchgang‘ (*first pass reading times*), Anzahl von Regressionen (Rücksprünge auf frühere Textpassagen etc.)

Beispiel

Daß die Frau der Mann störte, war unschön.

- **Vorteile:** Natürliches (unbeeinflusstes) Lesen; äußerst informativ, keine Tastenreaktion (schnelle Lesezeiten).
- **Nachteile:** Teuer, technisch aufwendig, schwierig auszuwerten.

Verarbeitungsverläufe (hypothetisch)



- On-line Methoden erlauben es, Verarbeitungsprozesse zeitlich aufzulösen (hier z.B. Trennung zwischen Abfolge- und Belebtheitseffekten)

"Visual World" (eye-tracking)



"The boy will **eat** the cake."

"The boy will **move** the cake."

- ⇒ Versuchspersonen bekommen visuelle Szene dargeboten, sowie (akustisch) eine darauf bezugnehmende linguistische Aussage.
- ⇒ *Vorwegnehmende* (anticipatory) Fixationen auf relevante Objekte (z.B. Kuchen), in Abhängigkeit vom linguistischen Input (z.B. „eat“; Altmann & Kamide, 1999).

Inferenzstatistik

- ⇒ Testen von **Hypothesen**, die aus Theorien und Modellen abgeleitet wurden.
- ⇒ Untersuchungen beruhen in der Regel auf VPn- und Item-**stichproben**. Inwieweit können Stichprobenergebnisse auf Sachverhalte in der jeweiligen Gesamtpopulation verallgemeinert werden?
- ⇒ Nutzung von inferenzstatistischen Verfahren, die die Stichprobengröße (und natürlich die in den Daten enthaltene Information) berücksichtigen.
- ⇒ Bestimmung der Wahrscheinlichkeit, mit der ein Unterschied (allgemeiner: ein Effekt) auf bloßen Zufall zurückzuführen ist (also lediglich eine Stichprobenschwankung darstellt).
- ⇒ Ist diese Wahrscheinlichkeit sehr klein, können wir die Annahme einer (Stichproben-)Zufallsschwankung verwerfen, und wir gehen von einem (auf die Population) verallgemeinerbaren Effekt aus.

Verfahrensschritte

1. Formuliere die Nullhypothese (H_0) („es besteht kein Unterschied“), und die dazugehörige(n) Alternativhypothese(n) (H_1) (H_2)...).
2. Wähle einen statistischen Test, um H_0 zu prüfen (hierbei spielen das Skalenniveau, bestimmte Verteilungsannahmen, sowie die Art des Versuchsplans eine Rolle).
3. Lege das Signifikanzniveau (α) und die Stichprobengröße (N) fest.
4. Finde die Stichprobenverteilung des Tests unter H_0 (s. Wahrscheinlichkeitstabellen).
5. Berechne die Prüfgröße des Tests mithilfe der Stichprobendaten. Ist die Wahrscheinlichkeit p des Auftretens der Prüfgröße kleiner oder gleich α , so lautet die Entscheidung, H_0 zurückzuweisen (\Rightarrow **signifikanter Unterschied**). Wenn nicht, so ist zu folgern, daß H_0 auf dem gewählten Signifikanzniveau nicht zurückgewiesen werden **kann**.

Beispiel

- ➔ Ein Mensch-Ärgere-Dich-Nicht Spieler hat den starken Verdacht, daß ein anderer Mitspieler einen manipulierten Würfel benutzt. Man beschließt diese Angelegenheit durch einen statistischen Test zu klären.
1. Formuliere H_0 :
Bei nicht-gezinktem Würfel: $p(1) = p(2) = p(3) = p(4) = p(5) = p(6)$
Alternativ dazu H_1 : Ungleiche Verteilung der Ereignisse.
 2. Wähle statistischen Test:
Skalenniveau? Art der Messung?
Nominal, unabhängige Messung
 \Rightarrow Chi-quadrat (χ^2) Test für eine unabhängige Stichprobe

3. Bestimme Signifikanzniveau und Stichprobengröße:
Man einigt sich auf $\alpha = 0.001$ und $N = 180$ (Würfe)

4. Finde die Stichprobenverteilung des Tests unter H_0 :
Schlag nach im Siegel (1985)

5. Berechne χ^2 unter Berücksichtigung der Daten:

$$\chi^2 = \sum_{(i=1-k)} (O_i - E_i)^2 / E_i$$

	k1	k2	k3	k4	k5	k6
O_{bserved}	16	18	30	36	40	40
E_{xpected}	30	30	30	30	30	30

$$\chi^2 = (16-30)^2/30 + (18-30)^2/30 + (30-30)^2/30 + (36-30)^2/30 + (40-30)^2/30 + (40-30)^2/30$$

$$= 6.53 + 4.8 + 0.0 + 0.2 + 3.33 + 3.33$$

$$= 18.19$$

6. Entscheidung:

Die Wahrscheinlichkeit p für $\chi^2 = 18.19$ unter H_0 ist zwar kleiner als 0.01, aber immer noch größer als α (0.001). Wir dürfen den Mitspieler deshalb nicht eines gezinkten Würfels bezichtigen (obwohl: Mißtrauen scheint durchaus angebracht).