

Vierte Sitzung: Experimentelle Methoden

Berry Claus

Computerlinguistik
Universität des Saarlandes

[Kontroversen in der Psycholinguistik]

Ist das Sprachverarbeitungssystem modular?

Ist Sprachverarbeitung regelbasiert?

Nachtrag von 3. Sitzung

VL Einführung in die Psycholinguistik SoSe 07

Berry Claus

4.1

[Modularität in der Sprachverarbeitung]

Konsens: Verschiedene Ebenen der Sprachverarbeitung
(Worterkennung, Parsing, Bestimmung von Bedeutung)

Dissens: Welche Beziehung besteht zwischen den Ebenen

Modul = in sich geschlossene Menge von Prozessen

Prozesse innerhalb eines Moduls arbeiten unabhängig von den
Prozessen anderer Module

Autonome Modelle

Sprachverarbeitung ist modular

Interaktive Modelle

Ebenen der Verarbeitung beeinflussen sich

Nachtrag von 3. Sitzung

VL Einführung in die Psycholinguistik SoSe 07

Berry Claus

4.2

[Fodors Modularität des Geistes]

These: Modulare Struktur des Geistes

bestimmte Aspekte von Kognition sind modular

→ Input-Systeme (Wahrnehmungsverarbeitung, Spracherkennung)

Charakteristika von Modulen

bereichsspezifisch

angeboren

verkapselt (*informational encapsulated*; informationsmäßig geschlossen:
kein Zugang zu modul-externen Informationen)

schnell

festverdrahtet (abgrenzbare neuronale Strukturen)

autonom

Nachtrag von 3. Sitzung

VL Einführung in die Psycholinguistik SoSe 07

Berry Claus

4.3

[Sprachverarbeitung = regelbasiert?]

Linguistisches Wissen: explizite linguistische Regeln

Satz kann aus einer Nominalphrase gefolgt von einer Verbalphrase bestehen

Pluralbildung im Englischen: Anfügen von -s (Ausnahmen)

Werden diese Regeln bei Sprachverstehen und -produktion verwendet?

Traditionelle Antwort: **JA!**

Konnektionistische Antwort: **NEIN!**

Konnektionistische Modelle: keine explizite Kodierung von Regeln
→ statistische Generalisierung

Beispiele: Graphem-Phonem-Korrespondenz, Erwerb von past-tense-Bildung

Nachtrag von 3. Sitzung

[Psycholinguistisches Experimentieren]

Psycholinguistik untersucht die mentalen Repräsentationen und Prozesse, die der menschlichen Sprachverarbeitung zu Grunde liegen

Empirischer Zugang: **Überprüfung** von Theorien und Modellen durch systematische Beobachtung menschlichen Verhaltens

Wichtigstes empirisches Instrument: Das **Experiment**
→ Theorien/Modelle werden in der Regel durch Experimente überprüft

[Experiment]

erlaubt die systematische und kontrollierte Beobachtung von Verhalten

ermöglicht so das Testen spezifischer Hypothesen

systematische Manipulation von Bedingungen:
→ Auswirkung auf Verhalten?

Unterschiede zwischen Bedingungen in Bezug auf Verhalten werden auf Manipulation zurückgeführt

Experiment erlaubt kausale Inferenz

Wichtig: Bedingungen dürfen sich nur hinsichtlich der experimentellen Manipulation unterscheiden

[Ein Beispiel]

Fragestellung: Ambige Nominalphrasen (z.B. Die Königin: Subj./Obj.) am Anfang eines Satzes → Wie werden die analysiert?

Unsere Theorie zur syntaktischen Analyse sagt: als Subjekt
Wie lässt sich das prüfen?

Wenn unsere Theorie stimmt, dann sollten sich bei Sätzen mit Obj.-Verb-Subj.-Abfolge Verarbeitungsschwierigkeiten zeigen

→ Hypothese: Längere Verarbeitungszeiten bei OVS als bei SVO

Testen im Experiment

Untersuchungsmaterial: Sätze in zwei Versionen: OVS und SVO:
Die Königin stört der Hofnarr. / Die Königin stört den Hofnarr.

Probanden erhalten die Hälfte der Sätze in OVS und die andere Hälfte in SVO (dazu später mehr)

[Noch ein Beispiel? – Das Erste]

Fragestellung: Verstehen Kinder nicht-chronologische Schilderungen? (Bsp.: Bevor Anna im Sandkasten gespielt hat, hat sie geschaukelt.) **Annahme:** Über-Fünffährige verstehen solche Sätze, jüngere Kinder nicht

Daten von Sprachtagebüchern

Die meisten Über-Fünffährige verstehen die chronologische Abfolge richtig, die meisten jüngeren nicht

Ist das ein Experiment?

NEIN

keine kontrollierten Bedingungen; inwieweit Unterschiede auf Zufall beruhen lässt sich nicht bestimmen; Unterschiede können andere Ursachen haben (z.B. im Beispiel: jüngere Kinder kennen die Bedeutung von *bevor* nicht); erlaubt nur sehr sehr schwache Konklusionen

[Noch ein Beispiel? – Das Zweite]

Selbe **Fragestellung** und **Annahme** wie bei „Das Erste“

Über- und Unter-Fünffährige Kinder im experimentellen Labor

→ zwei Gruppen von Kindern, die sich in anderen relevanten Variablen so weit möglich nicht unterscheiden; Kinder hören die nicht-chronologischen Sätze und sollen die genannten Ereignisse mit Puppen nachspielen; 2 Typen von Sätzen: nicht-chronologisch (*Bevor Anna ..., hat sie ...*) / chronologisch (*Anna hat ..., bevor sie ...*)

Ist das ein Experiment?

Nein, (aber) ein Quasi-Experiment

keine zufällige Zuordnung von Probanden zu Bedingungen, aber mehr Kontrolle; Konklusionen sind immer noch schwach bei manchen Fragestellungen ist kein anderes Vorgehen möglich (z.B. Geschlechtsunterschiede)

[Noch ein Beispiel? – Das Dritte]

Fragestellung: Bedeutung von Weltwissen für das Verstehen von Temporalsätzen bei Erwachsenen **Annahme:** Bei arbiträren Ereignisabfolgen (z.B. schaukeln → Sandkasten): Verarbeitungsschwierigkeiten bei nicht-chronolog. Schilderung; bei nicht-chronolog. geschilderten, nicht-arbiträren Abfolgen (Bsp.: *Bevor Anna den Apfel gegessen hat, hat sie ihn gewaschen*): keine Verarbeitungsschwierigkeiten

Untersuchung im Labor → Probanden lesen 4 Typen von Temporalsätzen (chronolog./nicht-chronolog. X arbiträr/nicht-arbiträr)

Ist das ein Experiment?

JA

systematische und kontrollierte Manipulation von Bedingungen; erlaubt stärkere Konklusionen (Alternativerklärungen können nie völlig ausgeschlossen werden)

[Unabhängige und Abhängige Variable]

Unabhängige Variable (UV)

experimentell manipulierte Variable (**Faktor**)

hat mindestens zwei Ausprägungen: Faktorstufen

häufig: Manipulation von mehr als einem Faktor (vgl. „Das Dritte“)

→ Testen von Interaktionen zwischen den Faktoren

experimentelle *Bedingungen*

bei einem Faktor: Bedingungen = Stufen; bei mehr als einem

Faktor: Bedingungen = Faktorstufenkombinationen

Abhängige Variable (AV)

erhobene/gemessene Variable

Operationalisierung der Größe, auf die laut Hypothese ein Einfluss der Manipulation der UV zu erwarten ist (bspw. Verarbeitungsschwierigkeiten: operationalisierbar durch Reaktionszeiten, Lesezeiten, Fehlerquoten, EEG)

[Versuchsplan]

Experimenten können verschiedene Versuchspläne zu Grunde liegen

Versuchsplan (Design) legt fest:

Anzahl der Faktoren

Anzahl der Stufen

z.B. 2 x 3 Design (zwei Faktoren, einer zwei-fach, der andere drei-fach gestuft)

und: Art der Zuordnung von Bedingungen zu Probanden (Pbn)

Zwei Möglichkeiten: Bedingungen können zwischen Pbn (*between-subjects design*) oder innerhalb von Pbn (*within-subject design*) manipuliert werden

[between-subjects design]

Aufteilung der Pbn in Gruppen: n Bedingungen → n Gruppen
jede der Gruppen wird eine der Bedingungen zugeordnet,
d.h. jeder Pb wird nur unter einer der Bedingungen getestet
(zwischen Pbn variiert Faktor: *Gruppierungsfaktor*)

Zuordnung von Pbn zu Gruppen/Bedingungen?

Randomisieren: zufällige Zuordnung

Ausgleichen: Zuordnung, so dass Gruppen einander in Bezug auf kritische Variablen möglichst entsprechen

Nachteil von between-subjects designs?

Gruppen und damit Bedingungen unterscheiden sich nicht nur in experimenteller Manipulation: systematische Unterschiede zwischen Pbn können nicht ausgeschlossen werden

aber: manchmal keine andere Möglichkeit als between-subjects design

[within-subject design]

jeder Pb wird unter jeder der manipulierten Bedingungen getestet
z.B. jeder Pb liest sowohl SVO- als auch OVS-Sätze
(innerhalb von Pbn manipulierter Faktor: *Meßwiederholungsfaktor*)

Vorteil von within-subject designs?

Bedingungen unterscheiden sich nicht in Bezug auf systematische Unterschiede zwischen den Pbn

Nachteil von within-subjects designs?

Übungseffekte, Ermüdungseffekte, Reihenfolgeeffekte

Nötig: möglichst verhindern, dass dadurch keine systematischen Unterschiede zwischen Bedingungen entstehen

[mixed design]

Kombination von between- und within-subjects design

z.B. 2 x 2 mixed design

Faktor 1 wird innerhalb Pbn variiert
(z.B.: OVS vs. SVO)

Meßwiederholungsfaktor

Faktor 2 wird zwischen Pbn variiert
(z.B.: visuelle vs. auditive Darbietung [Effekt von Intonation])

Gruppierungsfaktor

[Untersuchungsmaterial]

in psycholinguistischen Experimenten: Worte, Sätze, Texte
allgemein: *Items*, *Stimuli*

Manipulation von Bedingungen über Material

→ unterschiedliche items pro Bedingung

Problem?

wichtig: bis auf Bedingungsmanipulation sollten möglichst keine systematischen Unterschiede zwischen den items bestehen

Wie?

Versionen von items, die sich nur in Bedingungsmanipulation unterscheiden (z.B. OVS: *Die Königin stört der Hofnarr*; SVO: *Die Königin stört den Hofnarr*), bei n Bedingungen: n Versionen

nicht immer möglich, z.B. Experiment: hoch- vs. niedrigfrequenten Worten

generell: möglichst Ausgleich von items in Bezug auf Störvariablen (Länge, Plausibilität, Komplexität, Frequenz)

[Anzahl items]

Was wäre, wenn jeder Pb 1 Item pro Bedingung erhält

Ungenügend! zu wenig Daten, keine stabile Basis, Zufallseffekte
zudem: oft Datenverlust (Fehler, Extremwerte)

Wie viele Daten sind nötig? → Wie viele Items?

keine feste Regel, hängt auch von jeweiliger Untersuchung ab

Daumenregel: mindestens 5 Items pro Bedingung, besser mehr

[Zuordnung von Bedingungen zu Material]

within-subjects design: Jeder Pb wird unter jeder Bedingung getestet

Beispiel: OVS- versus SVO-Abfolge → 2 Bedingungen

Material: items in 2 Versionen (OVS/SVO); Anzahl Items: 20

→ insgesamt 40 Sätze (je 20 pro Version)

Liest nun jeder Pb alle 40 Sätze, also jedes Item in beiden Versionen?

Nein. Jeder Pb liest 20 Sätze, davon 10 in OVS und 10 in SVO

Wie wird zugeordnet, welcher Pb welche Items in welcher Version erhält?

Randomisieren: selten

Lateinisches Quadrat: häufig

[Zuordnung nach Lateinischem Quadrat]

Für unser Beispiel: Pbn werden zufällig in zwei Gruppen (PbnGr. 1 und 2) aufgeteilt, und items werden in zwei Gruppen aufgeteilt (ItemGr. 1 und 2)

Zuordnung von Items und Bedingungen zu Pbn

	OVS	SVO
PbnGr. 1	ItemGr. 1	ItemGr. 2
PbnGr. 2	ItemGr. 2	ItemGr. 1

→ jeder Pb liest jeweils die Hälfte der Items in OVS und die andere Hälfte in SVO; jedes Item wird von der Hälfte der Pbn in OVS und von der anderen Hälfte in SVO gelesen

Oft andere Beschreibung desselben Verfahrens: Zwei Listen von Items: Liste 1: Hälfte der Items in Bedingung 1 andere Hälfte der Items in Bedingung 2; Liste 2: umgekehrt

[Filler (Distraktor-Items)]

zusätzliche items im Experiment

haben keinen wirklichen Bezug zur eigentlichen Untersuchung

dienen zur Verschleierung der untersuchten Fragestellung
(Bsp: OVS/SVO: ohne Filler würden die Pbn ständig Sätze lesen, die mit einer femininen NP beginnen; Filler sollten dem "entgegensteuern": Sätze die mit maskulinen/neutralen NPs beginnen, teilweise SVO teilweise OVS)

Filler sollten keinen Bezug zu den Experimental-Items haben

[Ein Beispiel-Experiment]

Theorie 1

(Nur) bei ambigen satzinitialen NPs (z.B. *Die Königin ...*):
Verarbeitungspräferenz: analysiert als Subjekt

Theorie 2

Verarbeitungspräferenz (Analyse als Subjekt) auch bei nicht-ambigen Satzinitialen NPs (z.B. *Den König ...*)

Experimentelle Überprüfung?

Experiment, für dessen Ausgang die beiden Theorien unterschiedliche Vorhersagen machen würden

[Erste Operationalisierung]

Wie kann man messen, wie die satzinitiale NP analysiert wird?

Beispiel: ambige satzinitiale NP und OVS

Die Königin stört der Hofnarr.

Wenn NP1 (*Die Königin*) als Subjekt analysiert, dann sollte es bei NP2 (*der Hofnarr*) Verarbeitungsschwierigkeiten geben: Reanalyse

(keine Schwierigkeiten bei SVO)

Theorie 2: Auch nicht-ambige satzinitiale NP (Markierung als Dativ-Objekt durch Artikel) werden als Subjekt analysiert

→ Auch bei nicht-ambigen satzinitialen NP: Verarbeitungsschwierigkeiten bei OVS: *Den König stört der Hofnarr.*

[Design]

Welche / wie viele Faktoren, wie viele Stufen?

Ambiguität: 2-fach gestuft: ambig vs. nicht-ambig

Wortstellung: 2-fach gestuft: OVS vs. SVO

→ 2 x 2 Design

between-subjects oder within-subject? **within-subject!**

[Abhängige Variable und Methode]

Abhängige Variable?

Lesezeit für einen bestimmten Teil des Satzes

Methode?

Blickbewegungsmessung

[Materialkonstruktion]

Vier Versionen pro Experimental-Item

+A_OVS: *Die Königin stört der Hofnarr.*

+A_SVO: *Die Königin stört den Hofnarr.*

-A_OVS: *Den König stört der Hofnarr.*

-A_SVO: *Der König stört den Hofnarr.*

Kontrollieren

Plausibilität, Länge, Frequenz

Anzahl 32 Experimental-Items (→ 8 pro Bedingung)

Filler

[Zuordnung von Versionen/Items zu Pbn]

Pbn und Items werden in je vier Gruppen (PbnGr. 1 bis 4; ItemGr. 1 bis 4) aufgeteilt

Zuordnung von Items und Bedingungen zu Pbn

	+A_OVS	+A_SVO	-A_OVS	-A_SVO
PbnGr. 1	ItemGr. 1	ItemGr. 2	ItemGr. 3	ItemGr. 4
PbnGr. 2	ItemGr. 2	ItemGr. 3	ItemGr. 4	ItemGr. 1
PbnGr. 3	ItemGr. 3	ItemGr. 4	ItemGr. 1	ItemGr. 2
PbnGr. 4	ItemGr. 4	ItemGr. 1	ItemGr. 2	ItemGr. 3

[Darbietungsreihenfolge der Items]

festе und für alle Pbn gleiche Abfolge von Experimental-Items und Fillern (z.B. FFEFEFEFEFF)

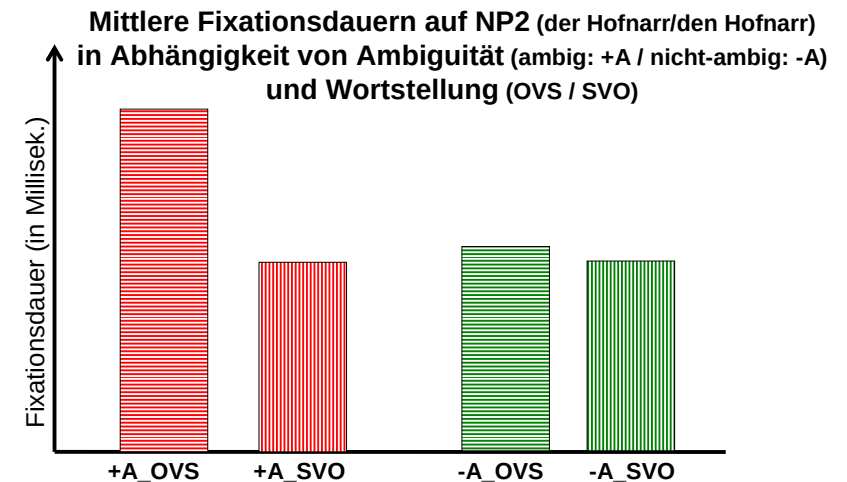
Innerhalb dieser festen Abfolge:

für jeden Pb unterschiedliche Darbietungsreihenfolge der Experimental-Items

[Welche Probanden?]

deutsch als Erstsprache
keine Sprachschwierigkeiten
keine visuelle Beeinträchtigung

[Daten (fiktiv!)]



[Und nun?]

Mittelwerte

Nur bei ambigen NP: Unterschied zwischen OVS und SVO

→ Evidenz für Theorie 1 und gegen Theorie 2?

Nein, da fehlt noch was

[Inferenzstatistik]

Angenommen: Mittelwerte zeigen den erwarteten Unterschied zwischen Bedingungen → empirische Unterstützung für Hypothese?

Nein! Experiment mit Stichprobe aus Population → vielleicht gibt es in der Population gar keine Unterschiede → Ergebnis kann zufällig zustande gekommen sein

Statistische Tests

testen gegen Nullhypothese (keine Unterschiede in Population)
ermitteln Wahrscheinlichkeit (p)

→ Wie hoch ist die Wahrscheinlichkeit dafür, Unterschiede in den experimentellen Bedingungen zu finden, wenn es in der Population keine Unterschiede gibt (= wenn Nullhypothese gilt)?

Wenn p klein ist, $p < .05$ / $p < .01$: Null Hypothese verwerfen

Effekt = **signifikant**

[Statistische Tests]

Unterschiedliche Tests

welchen man verwendet, hängt von verschiedenen Faktoren ab:
z.B. Skalenniveau, Verteilung der Messwerte

Ein knappes Beispiel: **Varianzanalyse**

testet auf Haupteffekte der einzelnen Faktoren und auf Interaktionseffekte (z.B. signifikante Interaktion zwischen Ambiguität von NP1 und Wortstellung: Wortstellung hat nur bei ambigen NP1 einen Effekt, nicht bei nicht-ambigen NP1)

kritischer Testwert, grob und verallgemeinert:

$$F = \frac{\text{Varianz aufgrund von experimenteller Manipulation}}{\text{Fehlervarianz}}$$

→ mehr zu statistischen Tests in der Sitzung am 20.06.

[Hausaufgabe]

Experiment zur Worterkennung entwerfen

Theorie 1

Worterkennung ist generell erleichtert, wenn zuvor ein semantisch assoziiertes Wort dargeboten wurde

Theorie 2

Nur bei semantisch ambigen Worten (wie *Bank*) ist die Worterkennung erleichtert, wenn zuvor ein semantisch assoziiertes Wort (wie *Geld* bzw. *Park*) dargeboten wurde

Hausaufgabe: Design, Methode, AV, Untersuchungsmaterial (ein Beispiel pro experimenteller Bedingung, Filler)

Hinweis: siehe Folien zu experimentellen Methoden von der letzten Sitzung