VL *Einführung in die Psycholinguistik*

Sechste Sitzung: Inferenzstatistik

Berry Claus

Computerlinguistik
Universität des Saarlandes

[Kohortenmodell]

Marslen-Wilson & Welsh (1978) **[nur AWE]**

Drei Verarbeitungsphasen

(1) **Zugriff**: Aktivierung lexikalischer Einträge auf Basis der perzeptuellen Repräsentation → Bildung einer Kandidatenmenge = **Kohorte**

(2) **Selektion**: Auswahl eines Elements der Kohorte

(3) **Integration**: „Nutzen" der syntaktischen/semantischen Eigenschaften des ausgewählten Wortes für Integration in Satzrepräsentation

Kohortenbildung → besonders relevant: Anfang eines Wortes
Wort-initiale Kohorte: große Menge von Kandidaten mit selben Anfang
Menge wird aufgrund weiterer phonologischer Information immer weiter eingeschränkt

[Kohortenmodell - Beispiel]

Input 1: /b/ initialer perzeptueller input
Kohorte 1: *Brief, Ball, Banane, Brot, Brombeere, Baum, brüllen ….*

Input 2: /br/
Kohorte 2: *Brief, ~~Ball, Banane~~, Brot, Brombeere, ~~Baum~~, brüllen ….*

Input 3: /bro/
Kohorte 3: *~~Brief~~, ~~Ball, Banane~~, Brot, Brombeere, ~~Baum~~, ~~brüllen~~ ….*

Input 4: /brom/
Kohorte 4: *~~Brief~~, ~~Ball, Banane~~, ~~Brot~~, Brombeere, ~~Baum~~, ~~brüllen~~*

Eliminieren von Kandidaten bis zum uniqueness point

neben phonologischer Information wird bei der Eliminierung **auch syntaktische und semantische Information** berücksichtigt

[Kohortenmodell – Revisionen und Bewertung]

Revisionen
Kontext kann nur die Integrationsphase beeinflussen (entsprechend experimenteller Befunde, wonach Kontext keinen Einfluss in frühen Phasen der Worterkennung hat)
Keine Alles-oder-Nichts-Eliminierung der Kandidaten: Ausmaß der Übereinstimmung ist bedeutsam

Notorisches Problem des Kohortenmodells
es gibt keine klaren Wortgrenzen → Schwierigkeiten, den für das Kohortenmodell wichtigen Anfang eines Wortes zu erkennen
Beispiel: *Die Kuh leckt das Kalb ab*. /Kuhle…/

Bis zum Hören von /ck/, ist unklar, ob es sich bei /l/ um einen Wortanfang handelt

[TRACE]

McClelland & Elman (1986) **[nur AWE]**

konnektionistisches Modell, stark interaktiv, betont Rolle von top-down Verarbeitung (Kontexteinflüsse) auf die auditive Worterkennung

besteht aus vielen simplen Verarbeitungseinheiten, verteilt auf 3 Ebenen:
(1) Einheiten für phonologische Merkmale → input level
(2) Einheiten für Phoneme
(3) Einheiten für Worte → output level

Einheiten der verschiedenen Ebenen, die **wechselseitig konsistent** sind, haben **exzitatorische Verbindungen**

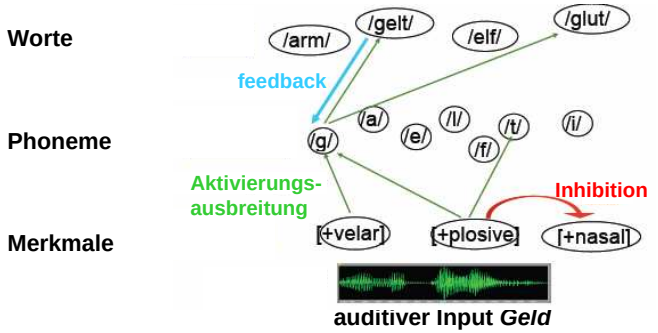
bidirektionale Verbindungen zwischen Ebenen: bottom-up und top-down

innerhalb einer Ebene gibt es inhibitorische Verbindungen

keine inhibitorische Verbindungen zwischen Ebenen

Effekt von lexikalisches Wissen bei ambigen input: z.B. /b/ vs. /p/ gefolgt von /ause/ → TRACE „erkennt" /p/"

[TRACE - bildhaft]



[TRACE - Bewertung]

implementiertes Computermodell: Vergleich von Simulationen des Modells mit menschlicher Sprachverarbeitung

gute Erfassung lexikalischer Kontexteffekte
gute Performanz beim Finden von Wortgrenzen
bewältigt verrauschten input

Hauptproblem
Annahme von starken Kontexteinflüssen (top-down-Effekte) auf die Worterkennung
Empirische Evidenz gegen diese Annahme: Klare Kontexteffekte lediglich bei schwer zu identifizierenden perzeptuellen input

[Heute]

Inferenzstatistik

[Wiederholung von 4. Sitzung: Wozu Inferenzstatistik?]

Inferenzstatistik → prüfen, ob das für untersuchte Stichprobe gefundene Ergebnis auf die Population (aus der die Stichprobe entnommen wurde) generalisierbar ist

Statistische Tests: testen gegen Nullhypothese (keine Unterschiede in Population) → ermitteln **Wahrscheinlichkeit (p)**

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

Signifikanter Effekt?
wenn $p < .05$ / $p < .01$, nennt man den Effekt der experimentellen Manipulation signifikant

[Wiederholung von 4. Sitzung: Wozu Inferenzstatistik?]

Was bedeutet **$p < .05$** inhaltlich?

→ Die Wahrscheinlichkeit dafür, Unterschiede zwischen den experimentellen Bedingungen zu finden, wenn in der Population die Nullhypothese gilt, ist kleiner als 5%

DANN Verwerfen der Nullhypothese ($=H_0$)
Akzeptanz der geprüften Hypothese ($=H_1$)

[Inferenzstatistik: schrittweise]

(1) Formulieren von H1 und H0 (meistens: „Unterschiede“ vs. „keine Unterschiede“)

(2) Bestimmen eines geeigneten statistischen Tests

(3) Festlegen von Signifikanzniveau (α) und Stichprobengröße (N)

(4) Datenerhebung

(5) Berechnen der Prüfgröße des statistischen Tests (auf Grundlage der bei der Stichprobe erhobenen Daten)

(6) Entscheiden, ob H_0 auf dem gewählten Signifikanzniveau zurückgewiesen werden kann oder nicht

[Geeigneter statistischer Test]

Wahl des statistischen Tests: abhängig von

Art der Hypothese
Unterschiedshypothese vs. Zusammenhangshypothese

Skalenniveau der erhobenen Daten (AV)
Nominal-, Ordinal-, Intervall- bzw. Verhältnisskala
qualitative Daten — quantitative Daten

weitere Merkmale der erhobenen Daten
Erfüllung der Voraussetzungen statistischer Tests

Aspekte des Designs
Anzahl Faktoren + Stufen, within-subject vs. between-subjects, Anzahl AV

[Parametrische Tests]

parametrische Tests: prüfen eine Hypothese über einen Parameter der Verteilungsfunktion der Grundgesamtheit (z.B. Mittelwertsunterschiede)

➔ **nötig**: Zufallsverteilung des Parameters unter Gültigkeit von H_0
Verteilungsfunktion muss bekannt sein

→ Stichprobenkennwerteverteilung = theoretische Verteilung beschreibt die Beziehung möglicher Ausprägungen eines statistischen Kennwertes und deren Auftretenswahrscheinlichkeit beim Ziehen von Zufallsstichproben des Umfangs n

➔ „typische“ parametrische Tests in der Psycholinguistik:
t-Test, **Varianzanalyse**

[Statistische Kennwerte (Stichprobe)]

Mittelwert: Arithmetisches Mittel
Quotient aus Summe aller Werte und Anzahl aller Werte $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Varianz: Maß zur Kennzeichnung der Variabilität einer Verteilung
Quotient aus der Summe der quadrierten Abweichungen aller Werte vom Mittelwert und der Anzahl aller Werte

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

Standardabweichung/Streuung:
Wurzel aus Varianz $s = \sqrt{s^2}$

[Statistische Kennwerte (theoretische Verteilung)]

Erwartungswert μ
→ Mittelwert der theoretischen Verteilung einer Zufallsvariablen

Varianz σ^2
→ Varianz der theoretischen Verteilung einer Zufallsvariablen

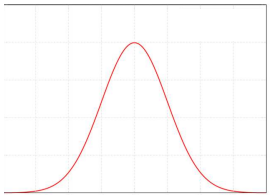
Schätzung der Populationsvarianz $\hat{\sigma}^2$
→ über Stichprobenvarianz
aber: Stichprobenvarianz unterschätzt Populationsvarianz um den Faktor (n-1)/n
deshalb:

$$\hat{\sigma}^2 = s^2 \cdot \frac{n}{n-1} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \cdot \frac{n}{n-1} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

Freiheitsgrade (df): Anzahl der Werte, die frei variieren können
hier: bei feststehendem Mittelwerte können alle bis auf einen frei variieren

[Normalverteilung (NV)]

Klasse von Verteilungen mit typischen Eigenschaften:
- glockenförmiger Verlauf
- symmetrisch um Erwartungswert
- Erwartungswert (→Mittelwert) =
- Modalwert (→häufigster Wert) =
- Medianwert (→halbiert Häufigkeitsverteilung)
- asymptotische Annäherung an x-Achse

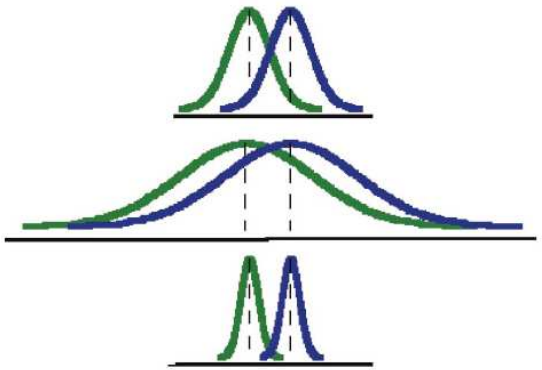


NV sind durch zwei Parameter eindeutig festgelegt:
- Erwartungswert und Streuung/Varianz

Bedeutsamkeit der NV: **Verteilungsmodell für statistische Kennwerte**
- Mittelwerte zufällig gezogener Stichproben = Zufallsvariable, die bei genügend großen Stichprobenumfang normalverteilt ist

Standardnormalverteilung
- Erwartungswert = 0 und Streuung = 1
- (durch z-Transformation lässt sich jede NV in eine StandardNV überführen)

[Rolle der Varianz]



mittlere Varianz

hohe Varianz

geringe Varianz

[Weitere Verteilungsfunktionen: t und F]

t-Verteilung und F-Verteilung

stehen in **Beziehung zu Standardnormalverteilung**
liegen parametrischen Test zu Grunde

t-Verteilung

Verteilung einer „komplexen Zufallsvariablen“: Quotient
im Zähler: normalverteilte Zufallsvariable; im Nenner: Quadrat einer normalverteilten Zufallsvariablen
liegt als theoretische Verteilung dem t-Test zu Grunde

F-Verteilung

Verteilung einer „komplexen Zufallsvariablen“: Quotient
im Zähler und Nenner: Quadrat einer normalverteilten Zufallsvariablen
liegt als theoretische Verteilung der Varianzanalyse zu Grunde

[t-Test für unabhängige Stichproben]

Vergleich zweier Stichprobenmittelwerte aus unabhängigen Stichproben

Indikation

Überprüfung einer Unterschiedshypothese für eine AV bei einem Faktor mit zwei Stufen, **between**-subjects manipuliert

Voraussetzungen

mindestens Intervallskalenniveau
bei kleinen Stichproben: Normalverteilung in Population
Varianzhomogenität

H_0: $\mu_1 - \mu_2 = 0$ → kein Unterschied der Stichprobenmittelwerte

H_1: $\mu_1 - \mu_2 \neq 0$ → unterschiedliche Stichprobenmittelwerte
(ungerichtete Hypothese)
(gerichtet, z.B.: $\mu_1 - \mu_2 > 0$)

[t-Test für unabh. Stichproben: Prüfgröße]

Prüfgröße = t

Verhältnis des Mittelwertunterschieds zur Streuung der Stichproben

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}}$$

wobei: $\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sum_{i=1}^{n_1}(x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2}(x_{i2} - \bar{x}_2)^2}{(n_1 - 1) + (n_2 - 1)} \cdot \left[\frac{1}{n_1} + \frac{1}{n_2}\right]}$

Standardfehler

Vergleich des empirisch ermittelten t-Wertes mit entsprechendem kritischen Wert der t-Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden und festgelegtem Signifikanzniveau

Empirischer t-Wert > krit. Wert → sign. Ergebnis

Realität: statistische Auswertungssoftware

[t-Test für unabh. Stichproben: Beispiel]

Frauen sind schnellere Worterkenner als Männer
AV: Reaktionszeiten bei lexikalischer Entscheidungsaufgabe

Daten

	Item 1	Item 2	…	Item 20	Mittelwert pro Person
Frau 1	630	650		600	
Frau 2	720	780		720	
…					
Frau 20	680	620		610	
				Mittelwert Frauen	$\bar{x}_1$
Mann 1	750	800		730	
Mann 2	1200	630		650	
…					
Mann 20					
				Mittelwert Männer	$\bar{x}_2$

$\hat{\sigma}_{\bar{x}_1}$ $\hat{\sigma}_{\bar{x}_2}$

[t-Test für abhängige Stichproben]

Vergleich zweier Stichprobenmittelwerte aus abhängigen (verbundenen) Stichproben

Indikation

Überprüfung einer Unterschiedshypothese für eine AV bei einem Faktor mit zwei Stufen, **within**-subjects manipuliert

Voraussetzungen

mindestens Intervallskalenniveau
bei kleinen Stichproben: Normalverteilung der Differenzen in Population

H_0: $\mu_d = 0$ → keine Mittelwertsdifferenz

H_1: $\mu_d \neq 0$ → Mittelwertsdifferenz unterscheidet sich von Null (ungerichtete Hypothese)
(gerichtet: $\mu_d > 0$)

[t-Test für abh. Stichproben: Prüfgröße]

Prüfgröße = t
Verhältnis des Mittelwerts der Differenzen zur Streuung der Differenzen

$$t = \frac{\bar{x}_d}{\hat{\sigma}_{\bar{x}_d}}$$ Standardfehler

wobei: $\bar{x}_d = \frac{\sum_{i=1}^{n} d_i}{n}$ und $d_i = x_{i1} - x_{i2}$

Vergleich des empirisch ermittelten t-Wertes mit entsprechendem kritischen Wert der t-Verteilung mit n – 1 Freiheitsgraden und festgelegtem Signifikanzniveau

Empirischer t-Wert > krit. Wert → sign. Ergebnis

Realität: statistische Auswertungssoftware

[t-Test für abh. Stichproben: Beispiel]

Schnellere Worterkennung bei semantischen Priming als bei Kontrollbedingung
AV: Reaktionszeiten bei lexikalischer Entscheidungsaufgabe

Daten

Pb 1	
sem 1	630
sem 2	650
....	
sem 10	710
control 1	715
control 2	800
...	
control 10	740
Pb 2	
...	
Pb 30	

	Mittelwert sem	Mittelwert control	Differenz der Mittelwerte
Pb 1			d_1
…			
Pb 30			d_{30}
Mittelwert der Differenzen			$\bar{x}_d$

[t-Test: Verletzungen der Voraussetzungen]

Sowohl bei abhängigen als auch bei unabhängigen Stichproben
→ t-Test reagiert robust auf Voraussetzungsverletzungen

Vermeiden bei unabhängigen Stichproben:
ungleichgroße Stichprobenumfänge

Problem bei abhängigen Stichproben:
negative Korrelation zwischen den Messwertreihen

[Varianzanalyse]

Indikation

Überprüfung einer Unterschiedshypothese für eine AV bei einem oder mehr Faktoren mit zwei oder mehr Stufen

Grundprinzip

Varianzanalyse zerlegt die vorhandene Varianz der Messwerte in Komponenten (→ **Quadratsummenzerlegung**, vgl. Summe der Abweichungsquadrate bei Varianzformel)

→ allgemein: (1) Varianzanteil der auf Manipulation des Faktors (= *treatment*) zurückzuführen ist

(2) restlicher Varianzanteil → Fehleranteil

[Varianzanalyse: Prüfgröße]

Prüfgröße

→ allgemein: F = zu prüfende Varianz / Prüfvarianz

→ Treatmentvarianz / Fehlervarianz

Vergleich des empirisch ermittelten F-Wertes mit entsprechendem kritischen Wert der F-Verteilung unter Berücksichtigung von Zähler-Freiheitsgraden und Nenner-Freiheitsgraden und festgelegtem Signifikanzniveau

Empirischer F-Wert > krit. Wert → sign. Ergebnis

Realität: statistische Auswertungssoftware

[Verschiedene Varianzanalysen (VA)]

je nach

Anzahl der Faktoren

einfaktorielle VA, zweifaktorielle VA, dreifaktorielle VA

Manipulation der Faktoren

within-subject → VA mit Messwiederholung
between-subjects → VA ohne Messwiederholung

je nach VA-Typ unterscheiden sich
zu prüfende Varianz und Prüfvarianz

[ohne vs. mit Messwiederholung: Beispiel]

Beispiel Fragestellung: Hängen die Lesezeiten für Sätze, in denen Ereignisse beschrieben werden, von der Dauer des beschriebenen Ereignisses ab?
→ Dreistufige Manipulation der Dauer des beschriebenen Ereignisses:
kurze / typische / lange Dauer (manipuliert über Durativadverbial)

im Prinzip zwei Möglichkeiten:
between-subjects → 3 Probandengruppen, die jeweils nur unter einer der drei Bedingungen getestet werden **➔ VA ohne Messwdhg.**
within-subject → jeder Proband wird unter jeder Bedingung getestet **➔ VA mit Messwdhg.**

[ohne vs. mit Messwdhg: Konsequenzen]

Sowohl bei VA ohne als auch bei VA mit Messwiederholung
zu prüfende Varianz: treatment-Varianz → Anteil der Varianz (Unterschiedlichkeit) in den Daten, der auf experimenteller Manipulation zurückzuführen ist

aber **unterschiedliche Prüfvarianzen**

[Prüfvarianz bei einfaktorieller VA ohne Messwdhg]

Gesamtvarianz setzt sich zusammen aus
Treatment-Varianz (Varianz zwischen Faktorstufen)
Fehlervarianz

Fehlervarianzanteil = Anteil der von experimenteller Manipulation unabhängig ist
→ Störvariablen: Messungenauigkeiten, Ausreißerwerte und **Unterschiede zwischen Probanden**

Prüfvarianz = Fehlervarianz

➔ Je größer die Unterschiede zwischen Probanden, desto größer die Fehlervarianz, desto kleiner der F-Wert

[Prüfvarianz bei einfaktorieller VA mit Messwdhg]

Gesamtvarianz setzt sich zusammen aus
Varianz zwischen Probanden
Varianz innerhalb von Probanden, die sich zusammensetzt aus
Treatment-Varianz (Varianz zwischen Faktorstufen)
Residual-Varianz (Interaktion (Pbn x Treatment) und Fehler)

Prüfvarianz = Residualvarianz

➔ Vorteil gegenüber VA ohne Messwiederholung
Varianz, die auf Unterschiede zwischen Probanden zurückzuführen ist, fließt nicht in die Berechnung der Prüfgröße (F-Wert) ein, Prüfvarianz ist geringer als bei VA ohne Messwiederholung