

Fünfte Sitzung: Worterkennung und Lexikalischer Zugriff

Berry Claus

Computerlinguistik
Universität des Saarlandes

[Planänderung!]

23.05. Worterkennung und Lexikalischer Zugriff

30.05. Inferenzstatistik und Computermodelle

06.06. Bedeutung und Textverstehen

13.06. Sprachproduktion

20.06. Syntaktische Verarbeitung

27.06. Situierete Sprachverstehen

04.07. Sprache und Embodiment

11.07. Klausurvorbereitende Wiederholung

18.07. Klausur



[Heute]

Worterkennung und Lexikalischer Zugriff

[Worterkennung und Lexikalischer Zugriff]

Welche Farbe hat das Wort auf der nächsten Folie?

BLAU

[Worterkennung und Lexikalischer Zugriff]

Welches Wort steht auf der nächsten Folie?

GELB

[Worterkennung und Lexikalischer Zugriff]

Das war der **Stroop-Effekt**:

Bei Farbbenennung: Interferenz, wenn Farbe und Inhalt des Wortes nicht übereinstimmen

Bei Wortbenennung: keine Interferenz

[Worterkennung vs. Lexikalischer Zugriff]

Worterkennung

Identifizierung des perzeptuellen (visuellen/auditiven) Inputs

Lexikalischer Zugriff

Suche und Auffinden eines Wortes im mentalen Lexikon auf Basis perzeptueller und kontextueller Informationen

[Mentales Lexikon]

beinhaltet die mentale Repräsentation des Wissens über Worte

enthält Informationen über **Formen** (Aussprache, Rechtschreibung) und **Bedeutung** von Worten

[Worterkennung + Lexikalischer Zugriff]

schnell (durchschnittliches mentales Lexikon: 50.000 bis 75.000 Worte)

automatisch

zuverlässig

[Auditive (AWE) vs. visuelle Worterkennung (VWE)]

auditiv

+ zeitlich gestreckt

+ flüchtig

- klare Wortgrenzen

visuell

-zeitlich gestreckt

-flüchtig

+klare Wortgrenzen

[Häufig verwendete Methoden (1)]

Lexikalische Entscheidungsaufgabe (lexical decision task) [VWE + AWE]

Präsentation einer Buchstabenabfolge [auch „auditive Variante“]

Probanden sollen per Tastendruck entscheiden, ob es sich dabei um ein Wort handelt oder nicht

AV: Reaktionszeit (Zeit zw. Präsentation des Wortes und Tastendruck)

Benennung (naming) [nur VWE]

Präsentation eines Wortes

Pbn sollen das Wort sagen

AV: Benennungslatenz (Zeit zw. Präsentation des Wortes und Sprechbeginn)

Priming-Paradigma

vor der Darbietung des **Zielwortes (Testwort, Target)** wird ein anders Wort (= **Prime**) dargeboten; **SOA** (stimulus-onset asynchrony): Zeit zw. Beginn der Präsentation des Primes und Beginn der Präsentation des Targets

[Häufig verwendete Methoden (2)]

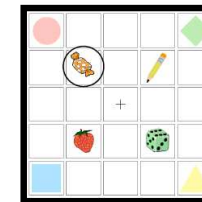
Blickbewegungsmessung während des Lesens [nur VWE]

AVs: Fixationsdauer (initial, gesamt), Regressionen

Blickbewegungsmessung beim Bildbetrachten [nur AWE]

Präsentation einer Abbildung von mehreren Objekten auf Monitor + gleichzeitige Instruktion eines der Objekte anzuklicken

AVs: Dauer bis zum Beginn einer Sakkade, Fixationshäufigkeiten



 *Klicke auf das Bonbon*

[Frequenzeffekte]

Hochfrequente (d.h., häufige) Worte werden schneller erkannt, und auf sie kann schneller reagiert werden als auf niedrigfrequente (d.h., seltene) Worte

„Monotoner Effekt“: nicht nur Unterschiede zwischen Extreme, sondern auch zwischen häufigen und etwas weniger häufigen Worten (z.B. *Regen* vs. *Pfütze*)

Korrelation zwischen Frequenz und anderen Variablen

Länge: Hochfrequente Worte sind oft kürzer

Age-of-acquisition (AOA, Erwerbsalter): Hochfrequente Worte werden früher gelernt; Kontroverse: Frequenzeffekte = AOA-Effekte?

[Semantisches Priming]

Worte werden bei vorheriger Präsentation eines bedeutungsähnlichen Wortes schneller erkannt (z.B. *Krankenhaus-Arzt* vs. *Kindergarten-Arzt*)

Priming = Methode → mit unterschiedlicher Wirkung: Prime kann Erkennen des Targets erleichtern oder erschweren

Assoziatives Priming vs. Nicht-assoziatives semantisches Priming

Zwei Worte sind **assoziativ verbunden**, wenn sie in einer Assoziationsaufgabe als Antwort genannt werden („Nenne das erste Wort, das Dir einfällt, wenn ich [Arzt] sage“)

Zwei Worte sind **nicht-assoziativ semantisch verbunden**, wenn sie in einer Assoziationsaufgabe zwar nicht genannt werden, aber dennoch bedeutungsverwandte sind (z.B. *Brot-Kuchen*, *Fuchs-Tier*)

Klare Evidenz für assoziatives Priming; unklare Befundlage für nicht-assoziatives semantisches Priming

[Wort/Nicht-Wort]

Bei lexikalischer Entscheidungsaufgabe wird auf Worte schneller reagiert als auf Nicht-Worte

Auf wenig plausible Nicht-Worte wird schneller reagiert als auf plausible

Langsame Reaktion auf ein Nicht-Wort wie *Zund*; schnelle Reaktion auf ein Nicht-Wort wie *Tnzvd*

Also: längere Reaktionszeiten bei **Pseudowörtern** = Nicht-Wort, dessen Zeichenfolge, den phonetischen und orthographischen Regeln der Sprache entspricht

[Uniqueness Point (nur bei AWE)]

Je eher sich ein Wort eindeutig identifizieren lässt, desto schneller die Antwort bei einer lexikalischen Entscheidungsaufgabe

Uniqueness point (Eindeutigkeitspunkt): Stelle der eindeutigen Identifikation

Beispiel: *Erdbeere* vs. *Brombeere*

Erdbeere (im Vgl. zu *Brombeere*): späterer uniqueness point

[Modelle der Worterkennung]

Suchmodelle vs. Modelle des direkten Zugriffs

Suchmodelle

perzeptuelle Repräsentation des sprachlichen inputs = Grundlage für eine Suche im mentalen Lexikon

→ Forsters Modell der autonomen seriellen Suche (VWE + AWE)

Modell des direkten Zugriffs

perzeptuelle Repräsentation: direkter Zugriff auf Einträge im mentalen Lexikon

→ Logogenmodell (VWE + AWE)

→ Interactive Activation and Competition Modell (nur VWE)

→ Kohortenmodell (nur AWE)

→ TRACE (nur AWE)

[Forsters Modell der autonomen seriellen Suche]

Forster (1976, 1979)

[VWE + AWE]

Bibliothekskatalogs-Metapher: (1) Suche im Katalog nach Buchstandort, (2) Gehe zum Standort, nehme das Buch + entnehme ihm Informationen

Forsters Modell: modalitätsspezifische **access files** (visuell, auditiv, Produktion), die auf Einträge im **master file** (=eigentliches mentales Lexikon) verweisen

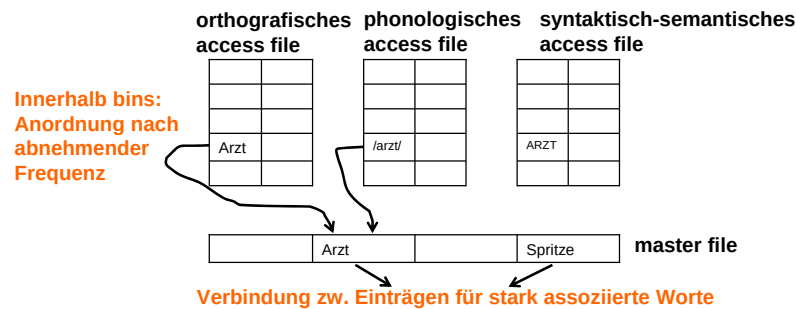
auf Basis der perzeptuellen Repräsentation des sprachlichen inputs wird das jeweilige access file seriell durchsucht

wird eine Übereinstimmung gefunden, dann wird der korrespondierende Eintrag im mentalen Lexikon aktiviert

nicht das gesamte access file wird durchsucht; access files sind unterteilt in **bins** (evtl. Einteilung nach erster Silbe bzw. erste Buchstaben)

→ Suche = eingeschränkt auf das jeweilige bin

[Forsters Modell – bildhaft]



[Forsters Modell – Bewertung]

autonom: kein Austausch zwischen Ebenen; einzig möglicher Kontexteffekt auf lexikal. Zugriff: assoziatives Priming im master file

gegen serielle Suche spricht: schnelle Worterkennung

andererseits: Modell liefert Erklärung für wesentliche Worterkennungseffekte

Probleme, Verarbeitung von Nichtwörtern zu erklären

[Logogenmodell]

Morton (1979)

[VWE + AWE]

Für jedes Wort im mentalen Lexikon gibt es zwei korrespondierende Logogen (eines für gesprochene und eines für geschriebene Sprache)

Logogen = **Merkmalszähler**

Zählerstand eines Logogens steigt, wenn perzeptueller input ein Merkmal des korrespondierenden Wortes enthält, Bsp.: visuelle Worterkennung: vertikaler Strich am Wortanfang → Aktivierung aller Logogene von Wörtern, die mit E, F, R, B ... beginnen

jedes Logogen hat einen spezifischen **Schwellenwert**, Bei Überschreiten des Schwellenwerts „**feuert**“ das Logogen → Input wird mit korrespondierendem Wort identifiziert

Nach Wortidentifikation: alle Merkmalszählerstände gehen in Ausgangszustand zurück

[Logogenmodell – Frequenz und Kontext]

Unterschiedl. Schwellenwerte für hoch- und niedrigfrequente Worte

Schwellenwerte von hochfrequenten Worten sind niedriger als die von niedrigfrequenten ← jedesmal, wenn ein Logogen feuert, wird sein Schwellenwert herabgesetzt

Erhöhung der Zählerstände durch Kontext

Logogensystems leitet seinen output an kognitives System weiter → Konstruktion einer Kontextrepräsentation → Vorhersagen, mit welchen Worten Äußerung wahrscheinlich fortgesetzt wird → Zählerstände der korrespondierenden Logogene werden erhöht

[Logogenmodell -- Bewertung]

Logogenmodell sagt Interaktion zwischen Kontext und Stimulusqualität und eine vergleichbare Interaktion zwischen Frequenz und Stimulusqualität vorher

Empirische Evidenz: Interaktion zwischen Kontext und Stimulusqualität ist größer und robuster als die Interaktion zwischen Frequenz und Stimulusqualität

Vorreiter/Wegbereiter konnektionistischer Modelle

[Interactive Activation and Competition Modell]

McClelland & Rumelhart (1981)

[nur VWE]

IAC Modell: Interaktives Aktivierungs- und Wettbewerbsmodell

konnektionistisches Modell

besteht aus vielen simplen Verarbeitungseinheiten, verteilt auf 3 Ebenen:

- (1) Einheiten für visuelle Merkmale → input level
- (2) Einheiten für Buchstaben
- (3) Einheiten für Worte → output level

Vernetzung der Einheiten

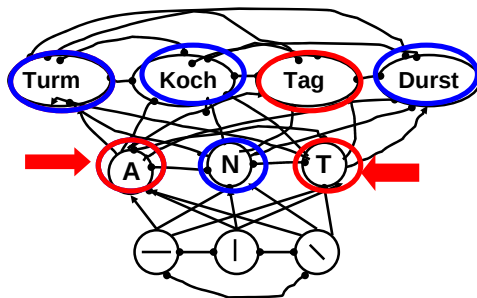
exzitatorische (positive) und **inhibitorische** (negative) **Verbindungen**

Beispiel: Buchstabe T

exzitatorische Verbindungen zu Einheiten für Worte mit T (→ Aktivierung) und inhibitorische Verbindungen zu Einheiten für Worte ohne T (→ Deaktivierung)

Einheiten innerhalb einer Ebene sind ausschließlich durch inhibitorische Verbindungen miteinander verbunden

[IAC Modell – bildhaft]



[IAC Modell - Bewertung]

kann Wortüberlegenheitseffekt erklären

stark eingegrenzter Geltungsbereich, z.B. keine Erklärung der Rolle von Bedeutung für Worterkennung

“per Hand“-Kodierung der Verbindungsstärken (also kein Lernen der Verbindungsstärken)

aber: sehr einflussreich

[nächste Sitzung]

Kohortenmodell und TRACE

[Lexikalische Ambiguität]

Werden bei der Verarbeitung lexikalisch ambiger (mehrdeutiger) Worte, wie Bank, Strom, Futter, alle Lesarten aktiviert?

Und was passiert, wenn der Satzkontext eine der Lesarten nahelegt?

Beispiel

Erschöpft setzte sich Hans auf die Bank an der Bushaltestelle. Da sah er, dass ein Geldtransporter vor der gegenüberliegenden Bank hielt.

[Experiment von Swinney (1979)]

Methode Crossmodales Priming mit lexikalischer Entscheidungsaufgabe
→ Pbn hören über Kopfhörer Sätze; an einem bestimmten Zeitpunkt während der Satzpräsentation (=kritische Stelle im Satz) wird auf einem Monitor eine Buchstabenfolge präsentiert → Wort vs. Nicht-Wort

→ Prime = auditiv präsentierter Satz, Target = visuell präsentiertes Wort

In Swinneys Experiment

Satz enthielt ein ambiges Wort, z.B. *bug* (Wanze)

Zwei Versionen des Satzes: neutraler Kontext vs. Kontext, der eine Lesart des ambigen Wortes nahelegt, z.B. Wanze = Insekt

Variation des Targets:

assoziiert mit nahe gelegter Lesart: *ant* (Ameise)

assoziiert mit der anderen Lesart: *spy* (Spion)

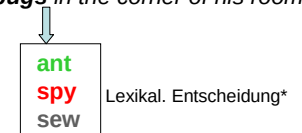
neutrales Kontrollwort: *sew* (nähen)

Zeitpunkt der Lexikalischen Entscheidungsaufgabe: unmittelbar nach ambigen Wort oder drei Silben später

[Swinney (1979) – früher Zeitpunkt]

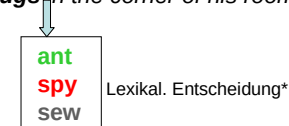
neutraler Kontext

*The man was not surprised when he found several **bugs** in the corner of his room.*



Lesart nahe legender Kontext

*The man was not surprised when he found several spiders, roaches, and other **bugs** in the corner of his room.*



Ergebnis für die Reaktionszeiten (RZ)

bei beiden Kontextbedingungen

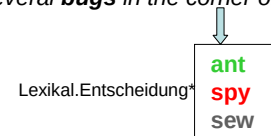
$RZ[\text{ant}] = RZ[\text{spy}] < RZ[\text{sew}]$

*es wurde jeweils nur eines der drei potentiellen targets präsentiert

[Swinney (1979) – später Zeitpunkt]

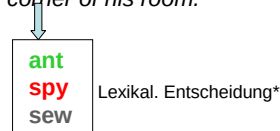
neutraler Kontext

The man was not surprised when he found several **bugs** in the corner of his room.



Lesart nahe legender Kontext

The man was not surprised when he found several spiders, roaches, and other **bugs** in the corner of his room.



Ergebnis für die Reaktionszeiten

neutraler Kontext: $RZ[\text{ant}] = RZ[\text{spy}] < RZ[\text{sew}]$

Lesart nahe legender Kontext: $RZ[\text{ant}] < RZ[\text{spy}] = RZ[\text{sew}]$

*es wurde jeweils nur eines der drei potentiellen targets präsentiert

[Swinney (1979) – Schlussfolgerung]

Konfrontation mit einem ambigen Wort → automatischer Zugriff auf alle Lesarten

Erst danach: Nutzung von Kontext → schnelle Entscheidung zwischen Lesarten → nur die kontextkonsistente Lesart bleibt aktiv

Semantischer Kontext kann nicht in den initialen Zugriff eingreifen