

Computational Psycholinguistics

Nutzung kognitiver Ressourcen beim
Satzverstehen: Die *Dependency Locality Theory*
(DLT, Gibson, 1998; 2001)

Überblick

- ⇒ Versuch einer Erklärung des Verhältnisses zwischen Sprachverarbeitungsmechanismen und zugrundeliegenden (*begrenzten!*) kognitiven Ressourcen.
- ⇒ Verarbeitungsschwierigkeit bestimmt durch zwei Komponenten:
 - **Integration Cost** (~ begrenzter Buffer)
 - **Storage Cost** (~ begrenzte Agenda)
- ⇒ Beide Komponenten bestimmen:
 - Den Aufwand bei der Verarbeitung nicht-ambiger Strukturen
 - Die Gewichtung (paralleler Parser!) von Alternativen im Falle von ambigen Strukturen.
- ⇒ Evidenzen zugunsten DLT
- ⇒ Mögliche Grenzen der DLT

Ziele / Ansprüche der Theorie

- ⇒ Erklärung der Verarbeitungsschwierigkeit (meßbar z.B. durch Komplexitätsratings bzw. on-line Lesezeiten) von **nicht-ambigen** Sätzen (siehe Zentraleinbettungen).
- ⇒ Erklärung der Gewichtung von Alternativen bei mehrdeutigen Sätzen (jedoch: Hier ist die Nutzung von Gedächtnisressourcen nur *ein* Faktor unter vielen, wie z.B. Plausibilität und lexikalischen Präferenzen).
- ⇒ Erklärung der Häufigkeitsverteilungen von Strukturen innerhalb einer Sprache (aufwendige Strukturen sollten weniger häufig auftreten).
- ⇒ Erklärung der Auftretenswahrscheinlichkeit bestimmter Kombinationen von syntaktischen Regeln/Mechanismen (im Sprachvergleich).
- ⇒ Erklärung von ‚Stufen‘ beim Spracherwerb.
- ⇒ Erklärung von Verstehensdefiziten bei Aphasikern.

Nicht-ambige Sätze: Zentraleinbettungen

- 1) The reporter¹ disliked the editor.
 - 2) The reporter¹ [_S who² the senator³ attacked] disliked the editor.
 - 3) ? The reporter¹ [_S who² the senator³ [_S who⁴ John⁵ met] attacked] disliked the editor.
- ⇒ Verarbeitungsschwierigkeit in (3) nicht bedingt durch Plausibilität oder lexikalische Häufigkeiten (=> begrenzte Gedächtnisressourcen):
- 3') John met the editor [_S who attacked the reporter [_S who disliked the editor]].

„Klassische“ Erklärungsansätze

- ⇒ Es gibt eine maximale Anzahl unvollständiger Abhängigkeiten, die gleichzeitig im Arbeitsspeicher aktiv gehalten werden können (z.B. Ingwe, 1960; Chomsky & Miller, 1963; Bever, 1970; Kimball, 1973; Abney & Johnson, 1991; Gibson, 1991; Lewis; 1993)
- ⇒ Die zitierten Modelle unterscheiden sich in der Art (und somit der max. Anzahl) von Abhängigkeiten, die Komplexitätsrelevant sein sollen, und ob Komplexität nur durch die Anzahl *gleicher* Abhängigkeiten bestimmt wird.

DLT: Ressourcen

- **Integration Cost** Kosten, die mit der Einbindung eines Wortes in die bereits (inkrementell) aufgebaute Satzstruktur verbunden sind;
Zentrale Idee: Die Kosten der Integration zweier Elemente (z.B. *Kopf* und *Argument* bzw. *Kopf* und *Adjunkt*; *Pronomen* und *Antezedens*) basieren auf **Locality**, d.h. Der Distanz zwischen beiden Elementen (wobei Distanz durch die Anzahl intervenierender Diskursreferenten operationalisiert wird):

z.B.: Subjekt-NP < ... mögliche intervenierende Diskursreferenten ... > Verb

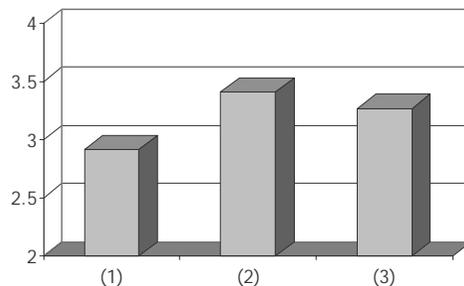
- **Storage Cost** (früher mal: **Prediction Cost**) Jeder lexikalische Kopf, der notwendig ist (sprich, prädiziert werden muß), um den gegenwärtigen Input-String zu einem grammatikalisch korrekten Satz zu vervollständigen, ist mit einer Storage Cost verbunden.

z.B. Subjekt ... (predict: Verb)
Objekt ... (predict: Subjekt & Verb)

Problem für ‚klassische‘ Ansätze: Eingebettete Pronomen

Warren & Gibson (1998)

- 1) The reporter [who the senator [who *I (you)* met] attacked] disliked the editor.
- 2) The reporter [who the senator [who *the professor* met] attacked] disliked the editor.
- 3) The reporter [who the senator [who *John* met] attacked] disliked the editor.



Complexity ratings for nested structures containing different kinds of NPs in the most embedded subject position. Scale from 1 (easy to understand) to 5 (hard to understand).

DLT: Integration Cost

- **Discourse Processing Cost (DPC)** Kosten verbunden mit der Konstruktion der Diskursstruktur der maximalen Projektion eines neuen Input-Wortes vom Typ *Kopf*. Vereinfacht: 1 energy unit (EU), wenn Kopf der Kopf eines **neuen** Diskursreferenten ist -- andernfalls 0 EUs.
- **Structural Integration Cost (SIC)** Kosten für die Anbindung einer Konstituenten an ihren bereits eingelesenen Attachment Host (hängt von der Anzahl neuer Diskursreferenten ab, die zwischen der einzubindenden Konstituente und dem Attachment Host Intervenieren). Vereinfacht: 1 EU für jeden intervenierenden neuen Diskursreferenten.
- **Schwerpunkt auf Diskurskomplexität**

Einfache vs. Doppelte Einbettung

Einfache Einbettung

	The	reporter	who	the	senator	attacked	disliked	the	editor.
DPC	0	1	0	0	1	1	1	0	1
SIC	0	0	0	0	0	2 ^a	2 ^b	0	0
total	0	1	0	0	1	3	3	0	1

^a Wegen Koindizierung einer Objektspur mit 'who' (2 neue Diskursreferenten dazwischen: 'senator' u. 'attacked')

^b Wegen 2 neuer Diskursreferenten ('senator, u. 'attacked') zwischen Subjekt ('the reporter') und Verb ('disliked').

Doppelte Einbettung

	The	reporter	who	the	senator	who	John	met	attacked	disliked	..
DPC	0	1	0	0	1	0	1	1	1	1	..
SIC	0	0	0	0	0	0	0	2	6 ^c	4 ^d	..
total	0	1	0	0	1	0	1	3	7	5	..

^c 2EUs für Integration von 'attacked' und 'the senator' ('John' u. 'met' intervenieren); 4EUs für Koindizierung von Objektspur mit 'who' ('senator', 'John', 'met', 'attacked' intervenieren)

^d 4EUs zur Integration von 'disliked' und 'the reporter' ('senator', 'John', 'met', 'attacked' intervenieren)

Eingebettetes Pronomen (/ oder *you*)

Doppelte Einbettung

	The	reporter	who	the	senator	who	YOU	met	attacked	disliked	..
DPC	0	1	0	0	1	0	0	1	1	1	..
SIC	0	0	0	0	0	0	0	2	4 ^a	3	..
total	0	1	0	0	1	0	0	3	5	4	..

^a 1EU für Integration von 'attacked' und 'the senator' ('met' interveniert); 3EUs für Koindizierung von Objektspur mit 'who' ('senator', 'met' und 'attacked' intervenieren)

On-line Effekte

- ⇒ Die (**off-line**) **Akzeptabilität** wird durch die **maximale** im Satz auftretende Total Integration Cost (bzw. Storage Cost) bestimmt.
- ⇒ Integration / Storage Cost sagt aber auch **on-line** Verarbeitungsaufwand vorher.

Beispiel: **Objekt-Relativsätze** (1) vs. **Subjekt-Relativsätze** (2):

(1) The reporter ₁ who the photographer ₁ sent to the editor ₃ hoped ₁ for a story. ₄

(2) The reporter ₁ who sent the photographer ₁ to the editor ₁ hoped ₄ for a story. ₁

DLT Vorhersagen (über Total Integration Cost):

- ⇒ **Objekt-RS**: 2 Maxima (*sent*, *hoped*), wobei *sent* < *hoped*
- ⇒ **Subjekt-RS**: 1 Maximum (*hoped*)

Objekt-Relativsätze

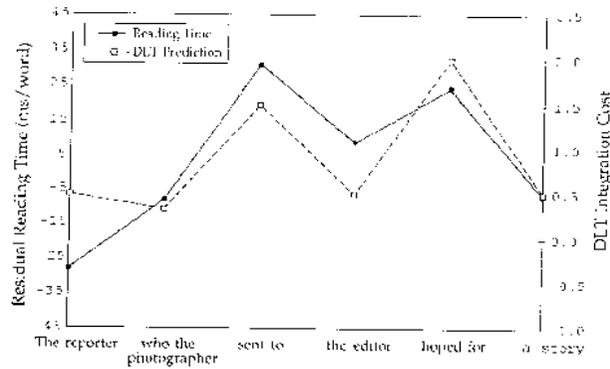
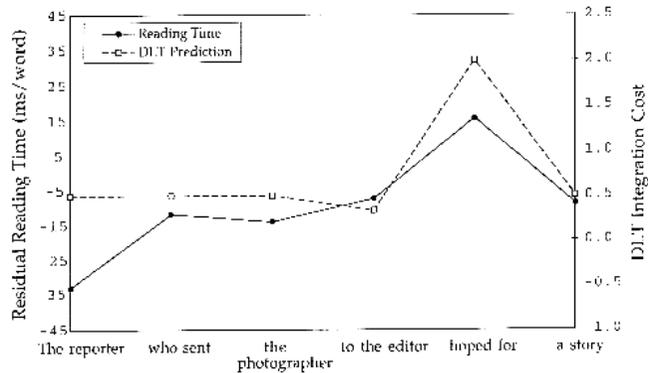


Figure 1: A comparison between residual reading times and locality based integration costs in an object-extracted RC structure.

- ⇒ Self paced reading Experiment von Gibson & Ko (1998): Gute Übereinstimmung zwischen RTs und Modellvorhersage.
- ? RTs für *the reporter*
- ? *sent* > *hope* in RTs (entgegen Vorhersage)

Subjekt-Relativsätze



Storage Cost

- ⇒ Jeder syntaktische Kopf der notwendig ist, um den gegenwärtigen Input-String zu einem grammatikalisch korrekten Satz zu vervollständigen, ist mit einer **Storage Cost** verbunden.
- ⇒ Vereinfacht: 1 memory unit (MU) für jeden lexikalischen Kopf, der vorhergesagt werden **muß** (minimale Prädiktion), um den gegenwärtigen Input zu vervollständigen.

Beispiel:

- (1) The reporter who the senator attacked disliked the editor.
2 1 2 3 2 1 1 1 0

(inkrementelle Vorhersage von Köpfen)

Auflösung von Ambiguitäten

- ⇒ Grundsätzlich gilt: Sowohl Total Integration Cost als auch Storage Cost (sowie andere Faktoren, wie lexikalische Präferenzen, Plausibilität usw.) bestimmen die initiale Präferenz.
- ⇒ Im Falle von Integration vs. Storage Cost Konflikten wird Storage Cost der Vorzug gegeben (da Storage Cost = potentielle Integration Cost).

Beispiel, in dem **Storage Cost** die initiale Präferenz bestimmen sollte:

(1) **Großer Storage Cost Unterschied:**

The witness who the evidence **examined** by the lawyer implicated was ill.

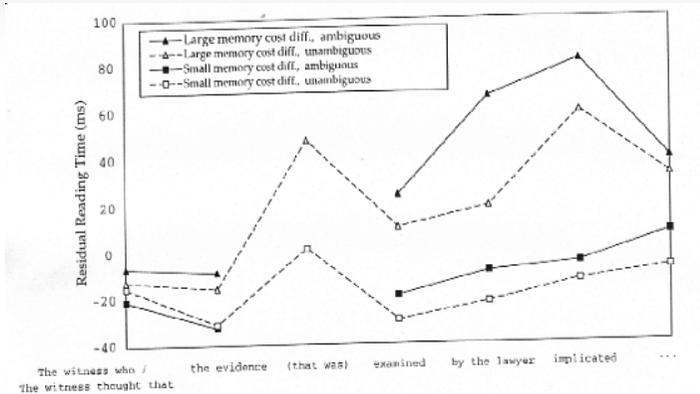
- ⇒ *examined* = main verb: 2 MUS (→ main verb, trace for *who*)
- ⇒ *examined* = participle: 3 (4?) MUS (→ main verb, RC-verb, trace for *who*, [ADV])

(2) **Kleiner Storage Cost Unterschied:**

The witness thought that the evidence **examined** by the lawyer implicated his wife.

- ⇒ *examined* = main verb: 1MU (→ object)
- ⇒ *examined* = participle: 1(2?) MU (→ main verb, [ADV])

Gibson, Grodner, & Tunstall (1997)



- ⇒ Low Storage Cost < High Storage Cost
- ⇒ Stärkerer Garden Path Effekt in High Storage Cost Bedingung
- ⇒ Storage Cost ist von Plausibilität und lex. Häufigkeit unabhängig

Übungsaufgabe

- ⇒ Gib für folgende Beispiele die Integration Costs und Storage Costs (im Satzverlauf) an:
 - (1) The [foster mother] who accepted the little boy [didn't really like] *herself* [at all] and [was reported] to the [social committee].
 - (2) The [foster mother] who accepted the little boy [didn't really like] *himself* [at all] [was reported] to the [social committee].