

Evidence for hierarchical sentence production

Zusammenfassung von Eric Auer <eric@coli.uni-sb.de>

31. August 2002

Dieser Text präsentiert die in meinem Seminarvortrag zum Seminar „Grammatical Encoding“ vorgestellten Resultate aus *Regulating Mental Energy: Performance Units in Language Production (K. Bock / C. Cutting 1992)* in Form einer kurzen Ausarbeitung – wer den Vortrag verpasst hat, kann also hier nochmal nachlesen worum es ging. Für Details, insbesondere genaue Statistiken zu den Messergebnissen von Bock und Cutting, verweise ich jedoch auf den Originalartikel (siehe oben), zu finden im *Journal of Memory and Language 31, Seiten 99 bis 127 (1992)*. Dort sind auch weitere Literaturverweise und die verwendeten Testsätze zu finden.

Inhaltsverzeichnis

1	Einleitung	2
1.1	Die Fragestellung:	2
1.2	Einheiten des syntaktischen Encoding:	2
1.3	Weitere Vorhersagen der beiden Modelle	3
2	Die Experimente	4
2.1	Experiment 1	4
2.2	Experiment 2	5
2.3	Experiment 3	5
3	Schlussfolgerungen	6

1 Einleitung

1.1 Die Fragestellung:

Eine alte Frage in der Psycholinguistik und Computerlinguistik ist die nach den Hintergründen entfernter Abhängigkeiten, sogenannter **Long Distance Dependencies**. Die Bock und Cutting untersuchen die Produktion von Sätzen, bei denen der Numerus zweier nicht direkt benachbarter Abschnitte voneinander abhängt. Dabei wird angenommen, dass die Satzproduktion entweder ein serieller oder ein hierarchischer Vorgang ist. Diese beiden Möglichkeiten werden anhand von drei Experimenten verglichen.

Die Methode: Die Experimente vergleichen die Vorhersagen die anhand eines seriellen und eines hierarchischen Produktionsmodelles über Schwierigkeiten bei der Satzproduktion und Satzvervollständigung gemacht werden können.

Das „serielle“ Modell verkettet nicht einfach Wörter linear, zum Beispiel anhand von Übergangswahrscheinlichkeiten oder N-gram Wahrscheinlichkeiten. Sätze mit entfernten Abhängigkeiten können nicht rein seriell Wort für Wort konstruiert werden, es muss noch einen **Stapelspeicher** geben, in dem für eine Weile Abhängigkeiten aufbewahrt werden können. Dieses Modell wird mehr belastet, wenn für lange Zeit viele Elemente auf dem Stapel bewahrt werden müssen.

Das andere Modell, welches von hierarchischer Satzproduktion ausgeht, hat mit anderen Satztypen Probleme. Insbesondere werden als neue Elemente **eingeschobene Klauseln** leichter verarbeitet, als wenn man die enthaltene Information in die aktuelle Phrase mit dazu packt, wodurch diese komplexer und schwerer zu verarbeiten wird. Die Experimente sollen herausfinden, welches Modell eher der menschlichen Sprachproduktion ähnelt. Eine typische psycholinguistische Fragestellung also.

1.2 Einheiten des syntaktischen Encoding:

Bei der Modellierung der menschlichen Satzproduktion ist das „**syntaktische Encoding**“ eine wichtige Komponente. Für die Übertragung per Sprache muss eine (semantische) Aussage erst in eine syntaktisch klare und korrekte Form gebracht werden. Dabei kommen verschiedene Größen für die angenommenen Satzbausteine in Frage.

Eine wichtige Einheit ist offenbar das einzelne **Wort**, mit kleineren Einheiten wie Silben oder Morphemen passieren dem Menschen kaum Fehler – Aussprachefehler passieren dagegen auf der Ebene von Phonemen, bei der

Umsetzung von Gedanken in gesprochene Sprache spielen diese also auch eine Rolle. Umgekehrt ist es aber auch wahrscheinlich, dass der Mensch Sätze unter Verwendung **grösserer** Einheiten konstruiert und nicht nur als blosser Aneinanderreihung von Worten in einem syntaktischen Grundgerüst. Wahrscheinliche Kandidaten für diese grösseren Einheiten sind syntaktische Sätze (und Phrasen) aber auch phonemische Sätze. Ein phonemischer Satz ist eine Wortfolge mit einer einzigen Betonung (-skurve).

Fehler bleiben oft innerhalb von Satzgrenzen, wie zum Beispiel im Versprecher „**Faster than the light of speed**“. Auch Pausen und Zögern und wie und wann Fehler bemerkt werden geben Hinweise auf die psychologische Bedeutung von Satzgrenzen. Sowohl finite als auch infinite Sätze scheinen dabei wichtig zu sein: Sowohl die bekannte Einheit „Verbform mit Argumenten“ als auch Ausdrücke wie „**Going home made Kathy happy**“ und „**Laura loves to program**“ scheinen Einheiten beim Satzbau zu sein. Hier werden jedoch nur finite Sätze untersucht.

1.3 Weitere Vorhersagen der beiden Modelle

Beim seriellen Modell mit Stapelspeicher ist die Arbeit beim Verarbeiten / Erstellen eines Satzes davon abhängig, wie sehr der Stapelspeicher belastet wird und wie lange. Auf dem Stapel werden insbesondere Forderungen zum Agreement (Übereinstimmung, zum Beispiel im Numerus) von Satzteilen gespeichert. Der Stapel ist damit von der Produktion des ersten bis zum Einbau des letzten Elementes aus einem Paar (oder einer Gruppe) voneinander abhängiger Satzteile mit einer „Agreement-Informationseinheit“ belastet.

Beim hierarchischen Modell gibt es eine **Funktionalen Stufe**, die Wortbedeutungen in (internen) Datenstrukturen verpackt und verarbeitet und eine **Positionalen Stufe**, die aus dieser strukturierten Darstellung einen linearen Satz, eine Wortkette, baut. Hier gehen Bock und Cutting davon aus, dass viel in derselben Datenstruktur verpackte Information diese schwierig zu verarbeiten macht. Dieselbe Information auf zwei Strukturen zu verteilen mache diese besser handhabbar. Für die Experimente wichtig ist dabei, dass eine zu komplexe Datenstruktur leichter zu Fehlern führt als zwei einfache Strukturen.

Da Bock und Cutting nach Beweisen für die grössere psychologische Realität eines hierarchischen Modells suchen, argumentieren sie: Dass Fehler leichter zum Plural hin passieren ist eine für das serielle Modell nur schwer zu begründende Asymmetrie. Beispiel eines solchen Fehlers: **... the result of the 1930s were...** – eine Erklärung mit dem seriellen Modell müsste annehmen, dass ein Plural auf dem Stapel „schwerer wiegt“ als ein Singular.

Ausserdem spricht es für das hierarchische Modell, dass zwischengeschobene Sätze einfacher als längere Sätze zu produzieren sind (wie das hierarchische Modell das erklärt, steht oben): **The claim that wolves were stealing babies was rejected** erscheint einfacher zu verarbeiten als **The claim about the stolen babies was rejected**, obwohl die zweite Fassung insgesamt kürzer ist – aber sie enthält mehr Information in einer einzigen Phrase. Die Information in der ersten Fassung ist dagegen übersichtlich auf zwei, wenn auch ineinander geschachtelte, Phrasen verteilt.

2 Die Experimente

Insgesamt präsentieren Bock und Cutting drei verschiedene Experimente zum **Number Agreement**. Sie messen jeweils die **Fehlerrate** in einer bewusst schwierigen Satzproduktionsaufgabe. Im ersten Experiment werden die Fehlerraten zwischen Präpositionalphrase und Relativsatz verglichen. Um Fehler durch eventuelle besondere Eigenschaften von Relativsätzen zu vermeiden bzw. den Verdacht einer Fehlmessung zu entkräften, wird im zweiten Experiment die Bedingung „Relativsatz“ durch eine Konstruktion mit einem NP (Nominalphrase) Komplement ersetzt. Im dritten Experiment dann wird der Einfluss der Länge von eingeschobenen Phrasen untersucht: Hier werden Präpositionalphrasen mit Komplementen bzw. Relativsätzen verglichen, wobei das hier das hierarchische Modell verschieden starken Einfluss der Phrasenlänge vorhersagt. Zusätzlich wurde noch die sogenannte „Sprechspanne“ der Versuchspersonen als Mass für die Leistung des (sprachlichen) Kurzzeitgedächtnisses der jeweiligen Versuchsperson gemessen.

2.1 Experiment 1

Die Versuchspersonen mussten Satzanfänge fortsetzen, wobei die Sätze entweder eine Präpositionalphrase wie in „**The editor(s) of the history book(s)...**“ oder einen Relativsatz wie in „**The editor(s) who rejected the book(s)...**“ im vorgegebenen Teil enthielten. Die verschiedenen Möglichkeiten für die Vorgabe des Numerus wurden nach Singular / Plural und übereinstimmend / unterschiedlich eingeteilt. Mit diesen drei Variablen ergeben sich also acht Kategorien für die Satzanfänge.

Die Ergebnisse lassen sich kurz zusammengefasst so beschreiben: Nach einem Plural traten mehr Fehler auf. Nach nicht übereinstimmendem Numerus traten mehr Fehler auf. Ausserdem traten mehr Fehler nach Präpositionalphrasen auf. Insgesamt traten kaum Fehler nach Singular auf, das Messergebnis zeigt hier also eine starke Asymmetrie – ein Punkt für das hierarchische Mo-

dell. Auch die höhere Fehlerrate nach einer Präpositionalphrase lässt sich damit erklären, wenn man annimmt, dass ein Relativsatz viel eher ein getrenntes Element ist. Siehe die Diskussion oben, in der zuviel Komplexität in einem einzelnen Element als für das hierarchische Modell schwierig angenommen wird.

2.2 Experiment 2

Da das Relativpronomen auch in einer „Agreement“ Beziehung zum Kopf der Phrase steht, könnte die Verarbeitung von Relativsätzen durch diese „Auffrischung“ des korrekten Agreement am Beginn des Relativsatzes allgemein einfacher sein (auch nach dem seriellen Modell) und damit die Messung verzerren. Darum wird in Experiment 2 die Bedingung „Relativsatz“ durch ein Komplement ersetzt. Als Kopf kommen jetzt nur noch Singular-NP (Nominalphrasen) zum Einsatz, um weniger Kategorien von Testsätzen (genauer: Satzanfängen) zu haben. Mögliche Satzanfänge sind also zum Beispiel: „**The report of the destructive fire(s)...**“ und „**The report that they controlled the fire(s)...**“. Dass damit Messverzerrungen verhindert werden können erscheint mir allerdings zweifelhaft. Auch hier unterscheiden sich die beiden Konstruktionen sicher in mehreren Aspekten, die auf verschiedene Weise mit den Modellen zusammenhängen bzw. für den Menschen wichtig aber von den Modellen eventuell gar nicht erfasst werden.

Der veränderte „Versuchsaufbau“ zeigt zusätzliche Ergebnisse: Nach unterschiedlichem Numerus bzw. überhaupt nach Plural treten mehr Fehler auf. Aber nur bei der Plural-Bedingung gibt es mehr Fehler nach einer Präpositionalphrase als nach einem Komplement. Nach einem Komplement wurde öfter der Satzanfang falsch zitiert (die Versuchspersonen mussten den Satz ergänzen und dabei den Anfang auch nochmal sagen), während nach PPs häufiger nichtflektierte Ergänzungen registriert wurden. Ein Komplement scheint mehr als eine PP eine abgegrenzte Einheit zu bilden, aber keine völlig in sich abgeschlossene Einheit. Das serielle Modell könnte annehmen, dass Komplemente das Kurzzeitgedächtnis mehr belasten, um die Ergebnisse zu modellieren

2.3 Experiment 3

Um die Idee weiter zu untersuchen, dass Komplemente stärker abgegrenzte Einheiten als PPs bilden, wurde Experiment 3 speziell auf diesen Einfluss abgestimmt: Wenn ein Komplement ein **getrenntes Element** ist, müsste seine **Länge** für die Satzergänzungsaufgabe **ohne Bedeutung** sein, wenn man vom **hierarchischen Modell** ausgeht. Die Idee ist, dass beim hierar-

chischen Modell der Abstand vom Kopf zur Satzergänzung dann immer „ein Komplement“ ist, während das serielle Modell die Länge des Komplementes noch als Faktor für die Verarbeitungsschwierigkeit vorhersagt. Die PP werden dagegen hier als Ergänzung des Kopfteiles betrachtet, der dadurch mit längeren PPs stets komplexer wird, auch im hierarchischen Modell also schwieriger zu verarbeiten (während die Komplemente „eingeschoben“ aber nicht „Teil von“ sind). Für dieses Experiment wurden die Satzanfänge von Experiment 2 verwendet, aber „fire“ alternativ als längeres „forest fire“ formuliert. Damit gab es wieder zweimal vier, also acht, Kategorien.

Das Ergebnis spricht für das hierarchische Modell: Nur bei PPs traten nach der längeren Version („forest fire“) mehr Fehler als nach der kürzeren Version auf. Ausserdem gab es wieder kaum Fehler nach Singular und allgemein mehr Fehler nach PPs als nach Komplementen. Nach PPs gab es wieder mehr „unflektierte“ Antworten und nach Komplementen mehr Erinnerungsfehler. Nach der längeren Version gab es allgemein mehr Erinnerungsfehler, und diese zeigten auch Abhängigkeit von der gemessenen Sprechspanne.

3 Schlussfolgerungen

Alle Experimente liefern Bock und Cutting Ergebnisse, die schon beinahe verdächtig gut zu ihrer Annahme eines **hierarchischen Produktionsmodelles** passen. Zusätzlich lässt sich aber sagen, dass das Wiederholen (und wohl auch das Verstehen) des gebotenen Satzanfanges Eigenschaften eines inkrementellen / seriellen Prozesses zeigen. Vereinfacht: Wenn der Satzanfang zu lang wird, treten Probleme auf, ihn korrekt zu zitieren. Also ganz gewöhnliche Effekte die sich mit einem in Einheiten von Worten seriell arbeitenden Kurzzeitgedächtnis gut erklären lassen.

Die Autoren merken allerdings an, dass man beim (angenommenen) hierarchischen Planungsprozess normalerweise nicht die wörtlich-exakte Repräsentation des Satzanfanges verwendet, da sich dies anstrengender „anfühle“. **Verstehen** von Sätzen wiederum hat sowieso völlig andere Falltüren und Klippen, hier seien als Beispiel nur die insbesondere von Frazier beschriebenen „Garden Path“ Effekte genannt, Am Ende des Textes von Bock und Cutting wird ausserdem noch kurz die Verwendung von Klauseln (zusätzlich zu Phrasen) als Einheit in der menschlichen Sprachverarbeitung diskutiert, dafür verweise ich den interessierten Leser auf den **Originaltext**. Auch die hier völlig weggelassenen statistischen Analysen und quantitativen Messergebnisse der oben beschriebenen Experimente sind dort zu finden.