

# Entwicklung eines Sprachkorpus zur Bestimmung des Sprachverstehens bei variabler linguistischer Komplexität

Verena N. Uslar<sup>1</sup>, Thomas Brand<sup>1</sup>, Mirko Hanke<sup>2</sup>, Rebecca Carroll<sup>3</sup>, Esther Ruigendijk<sup>3</sup>,  
Cornelia Hamann<sup>2</sup>, Birger Kollmeier<sup>1</sup>

<sup>1</sup>Medizinische Physik, <sup>2</sup>Amerikanistik, <sup>3</sup>Niederlandistik, Carl-von-Ossietzky Universität 9-11, 26111 Oldenburg,

## Einleitung

Verschiedene Forschungsarbeiten [1,2] zeigen, dass Sprachverstehen in akustisch anspruchsvollen Situationen nicht allein vom sensorisch-akustischen Hörvermögen abhängt, sondern auch von kognitiven Fähigkeiten des Hörers (z.B. Kapazität des Arbeitsgedächtnis oder maximale Leistungsfähigkeit des Sprachperzeptions-Systems). Hier wird die Evaluation eines Sprachkorpus mit Sätzen abgestufter linguistischer Komplexität beschrieben. Ziel der Evaluation dieses neu entwickelten Materials war die Erstellung des Oldenburger Linguistisch und Audiologisch kontrollierten Satztestmaterials (OLAKS). Dieser neue Satztest soll im Weiteren dazu verwendet werden, um den Einfluss von Satzkomplexität auf das Sprachverstehen qualitativ und quantitativ zu untersuchen.

## Satzmaterial

Das neu entwickelte Satzmaterial OLAKS besteht aus sieben Satztypen mit unterschiedlicher linguistischer Komplexität. Zum einen sind in dem neuen Material „Verb-2“-Sätze enthalten (Verben an zweiter Satzposition). Zu den Verb-2-Sätzen gehören die folgenden drei Satztypen:

1. subjektinitiale Sätze (Subjekt-Verb-Objekt, SVO)  
„Der stille Postbote begrüßt den dicken Frisör“
2. objektinitiale Sätze (OVS)  
„Den dicken Frisör begrüßt der stille Postbote“
3. ambige objektinitiale Sätze (amb OVS)  
„Die dicke Frisörin begrüßt der stille Postbote“

Im Fall der ambigen Objektinitialsätze wird erst mit dem zweiten Artikel deutlich, dass das Objekt vorangestellt wurde. Die Gruppe der Sätze mit Relativsatzeinbettung enthält folgende vier Satztypen:

4. Subjektrelativsätze ohne Ambiguität (SR)  
„Der Angler, der die Lehrerinnen umarmt, lacht“
5. Objektrelativsätze ohne Ambiguität (OR)  
„Der Angler, den die Lehrerinnen umarmen, lacht“
6. Subjektrelativsätze mit Ambiguität (amb SR)  
„Die Bauern, die die König grüßen, lachen“
7. Objektrelativsätze mit Ambiguität (amb OR)  
„Die Bauern, die die Königin begrüßt, lachen“

Bei den ambigen Relativsätzen wird die Mehrdeutigkeit erst spät im Satz, bei der Verwendungen des ersten Verbs aufgelöst.

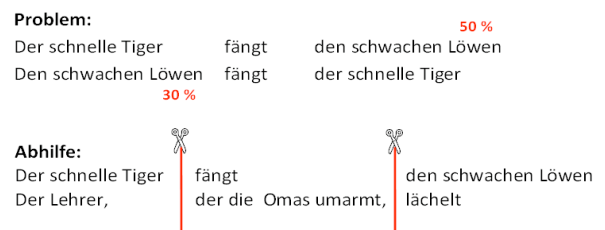
## Versuchsdesign

Die Evaluation des Materials gliederte sich in drei nacheinander stattfindende, aufeinander aufbauende Messphasen.

1. **Evaluationsphase I:** Messung mit Satzfragmenten von 720 Sätzen bei festem Signal-Rausch-Abstand (SNR: -7 dB)

2. **Evaluationsphase II:** Messung mit 560 verbliebenen ganzen Sätzen bei festem SNR (-7 dB)
3. **Evaluationsphase III:** Messung bei zwei weiteren satzspezifischer SNRs zur Bestimmung Satzdiskriminationsfunktionen mit 360 ausgewählten Sätzen

Entgegen der üblichen Vorgehensweise [3] wurde bei der Evaluation des neuen Materials zuerst die Verständlichkeit von Satzfragmenten bestimmt, da nur so sichergestellt werden konnte, dass die akustische Verständlichkeit zwischen den Bedingungen vergleichbar war. Das Beispiel in Abbildung 1 verdeutlicht die Problematik. Bei einer Messung mit ganzen Sätzen wäre nicht deutlich geworden, ob der Löwe in dem OVS-Satz schlechter verstanden wurde, weil er aus akustischen Gründen schlechter zu verstehen ist (schlechte Aufnahme, etc.), oder ob die Abweichung in der Verständlichkeit genau auf den syntaktischen Unterschieden beruht, die die Grundlage für spätere Messungen mit diesem Material bilden soll.



**Abbildung 1:** Messung mit Satzfragmenten in Evaluationsphase I um akustische und syntaktische Einflüsse besser trennen zu können. Die Verständlichkeit des Fragments „den schwachen Löwen“ sind in Prozent angegeben.

Nach jeder Messphase wurden aufgrund des Vergleichs mit den Mittelwerten und Standardabweichungen innerhalb einer Bedingung Sätze aussortiert, die nicht der in der jeweiligen Bedingung üblichen Verständlichkeit entsprachen, um später ein möglichst homogenes Material zu erhalten und den Messaufwand möglichst gering zu halten.

**Versuchspersonen:** An den Evaluationsmessungen nahmen insgesamt 36 normalhörende Probanden (12 pro Messphase) im Alter zwischen 20 und 30 Jahren teil. Alle Probanden waren Studenten der Carl von Ossietzky Universität Oldenburg.

**Versuchsaufbau:** Die Satzfragmente bzw. die ganzen Sätze wurden in einer schalldichten Kabine über Kopfhörer präsentiert. Nach der Wiedergabe eines Satzes/Satzfragments sollte die Versuchsperson alle gehörten Wörter des präsentierten Satzes wiederholen. Dabei war Raten explizit erlaubt. Der Versuchsleiter bestätigte jedes korrekt erkannte Wort auf einem Computerbildschirm, der für den Probanden nicht einsehbar war. Alle sieben Worte eines

Satzes gingen ohne Gewichtung in die Berechnung der Verständlichkeit mit ein (word-scoring).

## Ergebnisse

### Analyse der Unterschiede zwischen Satztypen

Abbildung 1 stellt die gemittelten korrekten Antworten der Messphasen 1 (jeweils die rechten Balken) und 2 (linke, farbige Balken) über alle Probanden und alle Sätze einer Bedingung und die entsprechenden Standardabweichungen dar. Sätze vom Satztyp 1 bis 3 sind bei Messungen mit ganzen Sätzen im Mittel zu 53 % zu verstehen, Sätze der Satztypen 4 bis 7 im Mittel zu 67 %. Bei Messungen mit Satzfragmenten liegen die Mittelwerte bei 59 bzw. 58 % wobei die Satzfragmente in den Bedingung 2 und 3 besser zu verstehen sind als die ganzen Sätze, im Gegensatz zu allen anderen Bedingungen Satzfragmente, die der Form nach das Subjekt eines Satzes bilden können, waren insgesamt besser zu verstehen, als Fragmente, die eindeutig das Objekt des Satzes waren (z.B. „die nette Lehrerin“ vs. „den netten Lehrer“).

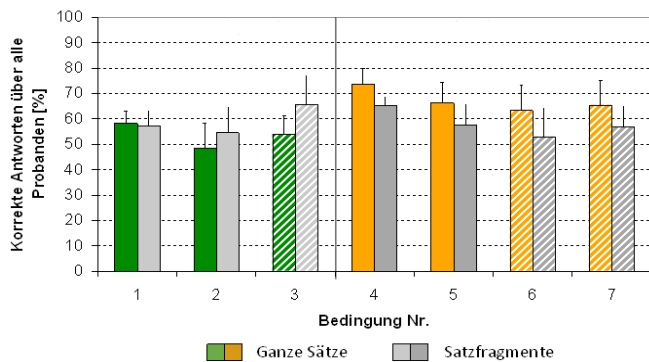


Abbildung 2: Korrekte Antworten gemittelt über alle Probanden getrennt nach Bedingungen und Evaluationsphase (Darbietungspegel: -7 dB SNR)

Die Satzdiskriminationsfunktionen für jeden der 360 Sätze wurden berechnet, indem mittels der drei Messwerte für jeden Satz bei unterschiedlichen SNRs die 50%-Sprachverständlichkeitsschwelle und die Steigung approximiert wurden. Abbildung 3 stellt die mittleren Sprachverständlichkeitsfunktionen (Verständlichkeit in Abhängigkeit vom SNR) für die sieben Satztypen dar. Auch hier zeigt sich der Trend, dass Satztyp 2 am schlechtesten zu verstehen ist und Satztyp 4 am einfachsten zu sein scheint. Eine detaillierte statistische Auswertung muss noch folgen.

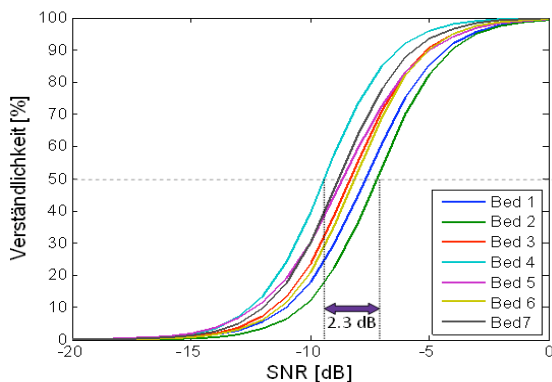


Abbildung 3: Satzdiskriminationsfunktionen (Verständlichkeit in Abhängigkeit vom Signal-Rausch-Verhältnis) gemittelt alle Sätze der jeweiligen Bedingung

### Analyse der Unterschiede zwischen Probanden

Abbildung 4 zeigt die korrekten Antworten pro Bedingung für sechs beispielhafte Probanden aus Messphase 3. Es sind deutliche Unterschiede in der Verständlichkeit zwischen den Probanden zu beobachten. Proband BL88 (blau) war in dieser Messung der beste Proband (70 % korrekt über alle Bedingungen) und Proband RG84 (rot) schnitt am schlechtesten ab (48 %). Allgemein lässt sich beobachten, dass ein Proband, der in einer Bedingung besonders gut oder schlecht abschnitt, dieses Verhalten auch in den anderen Bedingungen zeigte.

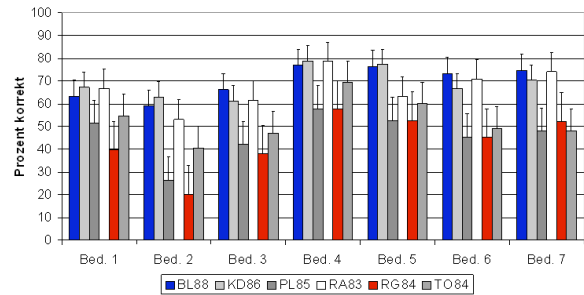


Abbildung 4: Korrekte Antworten über die Bed. für sechs der 12 Probanden aus Evaluationsphase 3.

## Zusammenfassung

Die Unterschiede in der Verständlichkeit zwischen den Bedingungen zeigen, dass das OLAKS-Material sehr gut dazu geeignet ist, den Einfluss linguistischer Komplexität auf das Sprachverstehen sowohl qualitativ als auch quantitativ zu untersuchen. Die schlechtere Verständlichkeit von Verb2-Sätzen im Vergleich zu den Sätzen mit Relativsatz-einbettung lässt sich mittels der durch die zusätzlichen Adjektive erhöhte Informationsdichte (und damit verbundene größere Anforderungen an das Arbeitsgedächtnis und andere kognitive Funktionen) erklären. Dass die Satzfragmente von Bed. 2 und 3 besser verstanden werden als die ganzen Sätze, liegt an der Satzstruktur der beiden Bedingungen. Nach dem Schneiden der Sätze aus Bedingung 3 erhält man zwei Fragmente, die der Form nach das Subjekt eines Satzes bilden können und sind somit besser verständlich als die Fragmente von Bedingungen 1 und 2, die jeweils nur ein eindeutiges Subjektfragment enthielten. Im ganzen Satz präsentiert zeigen sich bei den Bedingungen 2 und 3 dann die Effekte der ungebräuchlichen OVS-Satzstellung bzw. der Ambiguität. Die Verständlichkeit sinkt unter die Verständlichkeit für den Satztyp 1. Die großen individuellen Unterschiede deuten darauf hin, dass individuelle kognitive Fähigkeiten eine wichtige Rolle bei der Verarbeitung von Sprache spielen.

## Literatur

- [1] Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., Clarke Cox, L., Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *JAAA* 17 (2006), 487-497
- [2] Weik, V. N. et al.: Einfluss von Satzkomplexität auf das Sprachverstehen am Beispiel des Göttinger Satztests. 12. Jahrestagung der DGA (2009)
- [3] Wagener K.C., Brand T., Kollmeier B., Entwicklung und Evaluation eines Satztest für die deutsche Sprache I-III: Design, Optimierung und Evaluation des Oldenburger Satztest. *ZfA* 38 (1999), 4-15, 44-56, 86-95.