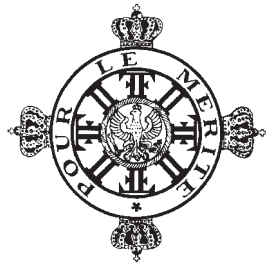

POUR LE MERITE

FESTVORTRAG

WILLEM J. M. LEVELT

ÜBER SPRACHTÄTIGKEIT – UNTERSUCHUNGEN ZUM SPRECHVORGANG

Sehr geehrter Herr Bundespräsident, sehr geehrter Herr Kanzler, sehr verehrte Damen und Herren,

die Lieblingsbeschäftigung aller Menschen ist wohl das Sprechen. Wir können uns stundenlang unterhalten, Geschichten erzählen, lehren oder uns streiten … und natürlich mit uns selbst reden. Gelegentlich wissen wir, *worüber* wir reden. Aber wir merken nie, *wie* wir reden. Wir nehmen kaum etwas von den hochkomplexen geistigen Prozessen wahr, die uns das Sprechen ermöglichen. Wir sind wie der Lachs, der das Wasser nicht kennt, in dem er schwimmt. Oft bewundern wir das erstaunliche Können von Berufsmusikern, Tänzern und Spitzensportlern. Aber eine noch viel erstaunlichere Fähigkeit, die wir alle besitzen, nehmen wir als selbstverständlich hin: das Sprechen kostet nichts.

Aber glücklicherweise gibt es Wissenschaftler für Sprache. Spätestens seit der Arbeit des indischen Sprachwissenschaftlers Panini, vor etwa zweieinhalbtausend Jahren, sind sie damit beschäftigt, das scheinbar Offenkundige zu ergründen und dabei nach und nach die Mysterien unserer Sprachfähigkeit zu lüften. Sie sind die Astrono-

men des menschlichen Geistes. *Ein* Sprachwissenschaftler soll besonders hervorgehoben werden, wenn ich hier in Berlin über den Sprechvorgang vortrage: Wilhelm von Humboldt (siehe Abbildung 1).
Die Sprachwissenschaftler seiner Zeit hatten begonnen, die faszinierenden Gemeinsamkeiten in den persischen, indischen und europäischen Sprachen, also in der indoeuropäischen Sprachfamilie, zu erkunden. Ihre Erkenntnisse beruhten ausschließlich auf der Analyse und dem Vergleich alter oder sehr alter geschriebener Texte. Humboldt erkannte, daß diese Texte Erzeugnisse der Sprache sind, aber nicht die Sprache an sich. Was aber ist dann Sprache?

> »Sie selbst ist kein Werk (Ergon), sondern eine Tätigkeit (Energeia). Ihre wahre Definition kann daher nur eine genetische seyn. Sie ist nämlich die ewig wiederholende Arbeit des Geistes, den articulierten Laut zum Ausdruck des Gedanken fähig zu machen.«

Sprache ist eine geistige Tätigkeit. Sie ist das, was der Sprecher tut, wenn er Gedanken in Schallwellen umsetzt. Genau das ist heute unser Thema.
Diese genetische Definition der Sprache fand im Laufe des 19. Jahrhunderts stets breitere Anerkennung. Der Höhepunkt war Wilhelm Wundts Werk *Die Sprache* aus dem Jahre 1900. Ordensmitglied Wundt beschrieb in seiner Theorie im Detail, was der Sprecher tut. (Siehe Abbildung 2a).
In seinem Leipziger Labor arbeitete ein etwa fünfundzwanzigjähriger amerikanischer Student, James McKeen Cattell (Siehe Abbildung 2b). Er war der erste, der – in Millisekunden – den genauen Zeitverlauf davon bestimmte, wie wir Bilder, Farben, Zahlen, Wörter und Buchstaben benennen. In meinem neuen Buch können Sie mehr über die spannende Schöpfung dieser jungen Wissenschaft lesen. (Siehe Abbildung 3)
Also, wie sprechen wir? Lassen Sie uns zunächst einmal die Sprache bei der Arbeit beobachten. Eine Mitarbeiterin an meinem Institut hat mir von ihrem Italienurlaub erzählt. Ich nenne sie hier Evelyn.

Hier ist ein Ausschnitt der Audioaufnahme, die ich davon gemacht habe – hier natürlich nur als Text wiedergegeben.

> *Ich möchte etwas über unseren Italienurlaub erzählen. Wir sind im Herbst gewesen; an der Westküste von Italien in einem relativ unbekannten und touristisch noch nicht so entdeckten Gebiet. Es liegt nördlich von … ahm … Ancona. Es ist ein Weingebiet. Es ist auch ein Wandergebiet. Die Strände dort sind leider noch nicht so erschlossen. Die sind … ah dort ist sehr viel Industrie.*

Wie können wir Evelyns Sprachtätigkeit analysieren? Wilhelm von Humboldt hat eine sehr nützliche Unterscheidung eingeführt: Es gibt danach zwei Aspekte des Sprechens und folglich der Sprache. Er nannte sie innere und äußere Sprache.
Innere Sprache: von Absicht zu Wortwahl und Satzbau (»Gedanken«).
Äußere Sprache: von Satz- und Wortform zu Silbenreihung und Artikulation (»Laut«).
Humboldt hat diese Unterscheidung nicht gerade genau definiert, aber seine Anhänger, unter ihnen auch Wilhelm Wundt, erkannten hierin die Möglichkeit, zwei Komponenten in der Produktion von Sprache zu unterscheiden. Innere Sprache bezieht sich auf die psychologischen Prozesse, die von einer Sprechabsicht zu Wortwahl und Satzbau führen. Äußere Sprache ist die Tätigkeit, die die Sätze und ihre Wörter in eine Abfolge ausgesprochener Silben umsetzt. Und tatsächlich hat sich gezeigt, daß diese beiden Komponenten der Sprachtätigkeit recht unabhängig voneinander sind. So entwickeln sie sich im Spracherwerb von Kindern weitgehend unabhängig; sie unterscheiden sich in ihrer neuronalen Basis im Gehirn (über die ich heute nicht reden werde), und sie lassen sich auch bei der Produktion von Sprache gut trennen. Betrachten wir also zuerst Evelyns innere und dann ihre äußere Sprachtätigkeit. In beiden Fällen werde ich die Hauptkomponenten dieser Tätigkeiten bloßlegen. Abbildung 4 zeigt alle diese Komponenten, die ich im Folgenden nacheinander besprechen werde. Ich nenne dieses Diagramm den »Grundriß« des Sprachsystems.

Also, als erstes hat Evelyn sich überlegt, was sie mir von ihren Ferien erzählen will. Sie tut dies in einer geordneten Abfolge von Aussagen. Diese Auswahl und Anordnung des Inhalts, also dessen, *was* gesagt werden soll, nennen wir *konzeptionelle Vorbereitung*.

Dies ist der erste Teil unseres theoretischen Grundrisses (siehe Abbildung 4, oben links). Wie schnell plant Evelyn den Inhalt ihrer Schilderung? Und wieviel Zeit braucht sie dazu, eine elementare Aussage – eine Proposition – auszusprechen? Die Gelehrten sind sich hinsichtlich der Definition von Propositionen nicht recht einig, aber als Faustregel kann man sagen, daß jede Proposition um ein Vollverb herum organisiert ist. Also habe ich auf jedes Vollverb in Evelyns Schilderung einen Klicklaut gesetzt. In Abbildung 5a sind diese Klicklaute als kleine vertikale Striche unter dem laufenden Sprachsignal plaziert. Es zeigt sich, daß es über die ganze Strecke von 27 Sprechsekunden nur 7 Striche, d. h. Vollverben gibt, etwa eins alle vier Sekunden. Das Planen und Anordnen von Information für eine Äußerung ist offenbar eine ziemlich langsame Angelegenheit. Das gilt nicht nur für Evelyn, sondern für alle Sprecher. Komplizierte Sachverhalte auszudrücken erfordert viel Aufmerksamkeit.

Lassen Sie uns einen Aspekt dieser Planung etwas genauer betrachten. Wie legen wir die Reihenfolge der relevanten Information in der Äußerung fest? Bei Berichten und Erzählungen ist das leicht – da folgen wir einfach der chronologischen Reihenfolge. Aber die Reihenfolge wird zu einem echten Problem, wenn man zum Beispiel die eigene Wohnung oder andere zwei- oder dreidimensionale räumliche Gegebenheiten beschreiben soll. Man muß dann die mehrdimensionale räumliche Anordnung in eine lineare zeitliche Abfolge von Propositionen umsetzen, ohne dabei das Ganze aus den Augen zu verlieren. Wo soll man nun anfangen, wo enden? Wir nennen dies *das Linearisierungsproblem* des Sprechers.

Um es zu untersuchen, haben wir Versuchsteilnehmer gebeten, Muster wie in den Abbildungen 6a, 6b und 6c zu beschreiben. In welcher Reihenfolge benennen sie die bunten Punkte und ihre Verbin-

dungen? Wir fanden heraus, daß Sprecher dabei drei einfachen Regeln folgen.

Die erste Regel ist, den Verbindungen zu folgen. Keine vernünftige Versuchsperson benennt bei Abbildung 6a erst den grünen, dann den violetten und dann den roten Punkt. Statt dessen gehen alle Versuchsteilnehmer von Grün, zu Blau, zu Rot, zu Gelb, und schließlich zu Violett. In all unseren Experimenten haben wir keine einzige Ausnahme von diesem Verbindungsprinzip gefunden.

Aber was tun, wenn es keinen *nächsten* Punkt mehr gibt, also wenn ein Pfad zu Ende ist? Das geschieht zum Beispiel, wenn Sie bei Abbildung 6b gemäß der Verbindungsregel vom blauen Punkt zum grünen Knotenpunkt gegangen sind und dann weiter zum grauen und violetten Knotenpunkt und schließlich zum roten Punkt. Dann greift die zweite Regel, nämlich zum letzten unerledigten Knotenpunkt zurückzukehren. Das ist hier der violette Punkt. Von hier aus geht man dann zum grünen Punkt. Anschließend springt man zum vorletzten, grauen Knotenpunkt und fährt dann mit der Beschreibung des blauen Knotenpunkts fort usw. Wenn man sich einen Weg durch so ein komplexes Muster bahnt, dann legt man Knotenpunkte sozusagen auf einem »Stapel« im Gedächtnis ab. Wenn man das Ende eines Pfades erreicht hat, dann geht man zum obersten Knotenpunkt auf dem Stapel zurück und beschreibt von dort aus einen neuen, noch nicht beschriebenen Pfad zum Ausgang. So arbeitet man rekursiv weiter, bis man alle aufgestapelten Knotenpunkte abgearbeitet hat. Die Regel ist hier »last in – first out« – »als letzter hinein, als erster hinaus«. Auch dieses Stapelprinzip gilt beinahe uneingeschränkt; unsere Versuchspersonen haben es nur äußerst selten verletzt.

Nun ist ein Knotenpunkt naturgemäß ein Punkt, an dem man sich zwischen zwei oder mehr Pfaden entscheiden muß. Welchen Pfad wählt man nun als erstes? Wie würden Sie z. B. fortfahren, wenn Sie den blauen Knoten auf Abbildung 6c erreicht haben? Hier kommt die dritte Regel, eine Wahrscheinlichkeitsregel, zum Zuge: dieser Regel zu Folge wählt man zunächst den einfachsten Pfad. Dies ist der Pfad mit der geringsten Anzahl weiterer Knotenpunkte. Demzu-

folge wählten unsere Versuchspersonen hier am blauen Wahlpunkt also zumeist den rechten Pfad. So brauchten sie keine weiteren Knotenpunkte zu ihrem Stapel im Gedächtnis hinzuzufügen. Wenn sie nach links gegangen wären, dann hätte sich kurzfristig ein Stapel mit zwei Knotenpunkten, dem blauen und dem gelben Punkt, gebildet; und das haben die Versuchsteilnehmer vermieden.

Wir konnten zeigen, daß diese drei Regeln nicht nur bei der Linearisierung visueller Gegebenheiten Anwendung finden, sondern sehr allgemeiner Natur sind. Sie werden zum Beispiel auch angewandt, wenn man seine Verwandtschaftsbeziehungen beschreibt. Es gibt hier übrigens eine interessante Parallele zwischen dem Linearisierungsproblem des Sprechers und dem Problem, auf das Chemiker stießen, als sie die komplexen räumlichen Strukturen von Molekülen in linearen chemischen Formeln abbilden wollten.

Die Linearisierung setzt sich fort, während man spricht. Gehen wir zu der Beschreibung von Evelyn zurück. Evelyn hat also nicht schweigend die ganze Abfolge von Propositionen geplant, bevor sie mit der Schilderung anfing. Statt dessen hat sie zu sprechen angefangen, sobald der erste Gedanke zur Äußerung bereitstand. Mit Hilfe des Verbindungsprinzips konnte sie den nächsten Gedanken aus dem Gedächtnis abrufen, während sie den ersten sprachlich ausdrückte. Sprache wird stückchenweise, inkrementell, erzeugt. Wie wir sehen werden, gilt dies nicht nur für die Planung des Inhalts, sondern auf allen Ebenen der Sprachtätigkeit, einschließlich der Aussprache.

Natürlich muß der Inhalt jeder einzelnen Proposition im Detail geplant werden. Wenn Sie das Muster in Abbildung 6c beschreiben, dann können Sie erklären, daß man beim blauen Punkt links zum roten Punkt gehen muß. Sie fangen also an, die Äußerung »Nach links gehen zum roten Punkt« zu planen. Hierbei wählen Sie unter anderem ein Bewegungskonzept (GEHEN), ein Richtungskonzept (LINKS), eine Farbe (ROT) und eine Form (PUNKT) aus – siehe Abbildung 7. Wir nennen diese Begriffe lexikalische Konzepte, weil wir dafür Wörter in unserer Sprache, in unserem mentalen Lexikon, haben. Die Bedeutung eines Eintrags in unserem Lexikon ist ein

lexikalisches Konzept. Es ist rechts oben im Grundriß (Abbildung 4) abgebildet. Dieses blaue Oval stellt die Bedeutung des Lexikoneintrags dar: ROT oder PUNKT oder GEHEN oder LINKS.

Wir wissen aber mehr über ein Wort als nur seine Bedeutung. Wir wissen auch, wie es im Satz verwendet werden kann. Das ist die syntaktische Information über das Wort. Wir sprechen hier vom *Lemma* eines Wortes. Siehe das blaue Oval unter »Konzept« rechts oben in Abbildung 4. Wir wissen zum Beispiel, daß »gehen« ein intransitives Verb ist – es kann nicht direkt mit einem Objekt verbunden werden. Man kann zum Beispiel nicht sagen »ich gehe den Hund«. Oder nehmen Sie »Punkt«. Es ist ein Nomen, das mit einem Adjektiv, etwa »rot«, kombiniert werden kann.

Lemmata ähneln Legosteinen, die auf bestimmte Weise zusammengesetzt werden können. Wir können damit komplexe syntaktische Muster bauen. Die Lemmata ROT und PUNKT können sich verbinden und so die Phrase »roter Punkt« hervorbringen. Dies ähnelt ein wenig dem Aufbau komplexer chemischer Moleküle aus einer kleinen Gruppe von Atomen.

Schließlich gehören zu jedem Wort auch Angaben zur Aussprache. Jedes Wort hat eine bestimmte phonologische Form, die aus Vokalen, Konsonanten, d. h. Phonemen besteht. Wir bezeichnen dies als die Form des Wortes (siehe das braune Oval in Abbildung 4) Das Wort »rot«, zum Beispiel, hat als Wortform die Phonemfolge *r-o-t*.

Ein vollständiger Lexikoneintrag ist also eine Triade von Bedeutung, Syntax und Form. Als gebildete Erwachsene kennen wir etwa fünfzig- bis hunderttausend solcher Triaden. Das ist unser *mentales Lexikon*.

Gehen wir nun zurück zu unserer Sprecherin, Evelyn. Sie erzeugt einen Strom von Wörtern. Wie schnell geschieht dies? Ich habe in Abbildung 5b jedes Wort im Sprachsignal mit einem Strich markiert. Wie Sie sich erinnern, war das Tempo der Propositionen in Evelyns Schilderung recht gemächlich. Bei den Wörtern ist das anders. Sie haben hier eine Geschwindigkeit von etwa zwei Wörtern pro Sekunde, aber es können auch leicht bis zu vier oder fünf Wörter pro Sekunde werden. Das ist also das Tempo, mit dem Evelyn Wörter

aus ihrem Lexikon mit seinen 50.000 bis 100.000 Einträgen auswählt. Dieser Selektionsprozeß ist nicht nur rasend schnell, sondern auch sehr robust. Daß wir etwas Falsches auswählen, so wie hier, ist eher selten:

Lexikalische Selektionsfehler

Und dann kommst du zu *Blau* – ich meine Grün
… ich kann ihr nichts *Gutes* … ah! Schlechtes nachsagen.
… hast du mal einen *Radiergummi* [Gemeint: Spitzer]
… das Einschlagen eines *Hammers* [Gemeint: Nagels]
(aus den Bierwisch, Leuniger, Meringer und Levelt Corpora)

Die durchschnittliche Fehlerrate ist geringer als ein Fehler pro 1000 Wörter.
Zurück zum fehlerfreien Sprechen. Ganz wichtig ist, daß Wörter nicht in zufälliger Folge auftreten, sondern in einer wohlgeformten syntaktischen Abfolge stehen. Dieser Prozeß der Auswahl von Lemmata zur Äußerung und ihrer Anordnung in der richtigen syntaktischen Reihenfolge mitsamt den richtigen grammatischen Markierungen heißt »grammatische Kodierung«. Sie ist in Abbildung 4 links oben als zweites blaues Feld im Grundriß angedeutet. Die Abfolge der Wörter im Satz muß natürlich den syntaktischen Eigenschaften der ausgewählten Lexikoneinträge, ihren Lemmata, den Legosteinen, entsprechen. Ab und zu entstehen bei der grammatischen Kodierung syntaktische Fehler wie diese:

Syntaktische Fehler

Zu übersetzen gelungen ist mir der Brief nicht.
Dann muß daran versucht werden, etwas zu ändern
(aus dem Bierwisch Corpus)

Diese beiden Aspekte der grammatischen Kodierung, die Auswahl und die serielle Ordnung von Wörtern, haben wir ausführlich untersucht. Ich möchte zunächst kurz auf die Auswahl von Lemmata eingehen.

Bei der Auswahl von Lemmata für eine Äußerung entsteht immer eine Art Konkurrenz zwischen Wörtern mit ähnlicher Bedeutung. Deshalb sagen wir manchmal etwa *links* statt *rechts* oder *Frage* statt *Antwort*. Wie stark die Konkurrenz zwischen Wörtern ist, kann man messen, und das haben wir, buchstäblich in Dutzenden von Experimenten getan. Ich möchte Ihnen ein solches Experiment als Beispiel beschreiben. Sehen Sie sich einmal in Abbildung 8 die 25 Bilder aus einem Bildbenennungsexperiment an. In der ersten Reihe stehen Transportmittel: Zug, Fahrrad, Auto und so weiter. In der zweiten stehen Werkzeuge, zum Beispiel ein Besen und eine Axt, in der dritten Reihe stehen Tiere, in der vierten Möbel und in der fünften schließlich Kleidungsstücke. Im Experiment sind wir folgendermaßen vorgegangen: In jedem Versuchsdurchgang sah die Versuchsperson ein Bild und sollte es so schnell wie möglich benennen. Im ersten Testblock haben wir nur die Transportmittel aus der ersten Reihe benennen lassen, und zwar jedes Objekt mehrmals, in zufälliger Abfolge. Wir haben dabei die Zeit vom Bildbeginn bis zum Anfang des gesprochenen Wortes gemessen. Das sind im Mittel etwa 600 ms. Im nächsten Testblock haben wir dann nur die Werkzeuge aus der zweiten Reihe benennen lassen, wieder viele Male in zufälliger Reihenfolge, danach die Tiere und so weiter. Wir nennen dies die homogene Versuchsbedingung, weil hier alle Bilder, die im selben Testblock vorkommen, zur selben semantischen Klasse gehören.

Man kann aber auch spaltenweise testen. Dann ergibt sich bei der ersten Spalte: Zug, Besen, Maus und so weiter. Das haben wir auch nacheinander für jede der fünf Spalten, gemacht. Wir nennen dies die heterogene Bedingung, weil nun die Bilder, die im selben Testblock vorkommen, verschiedenen semantischen Klassen angehören. Nun kommt die experimentelle Frage: Nehmen Sie zum Beispiel die Maus. Was meinen Sie, wird die Maus nun in der homogenen oder in der heterogenen Bedingung schneller benannt? In der homoge-

nen Bedingung, wo man nur Tiere zu benennen hat, wissen Sie als Versuchsperson bereits, daß auch das nächste Bild wieder ein Tier sein wird. Man sollte meinen, daß Ihnen das hilft, schnell zu reagieren. Unsere Konkurrenztheorie sagt aber etwas anderes vorher. Achten Sie auf das Zielwort *Maus* (in Abbildung 8 umzirkelt). Abbildung 9a zeigt, was in der homogenen Testbedingung geschieht, wenn Sie die Maus anschauen, um sie zu benennen. Sie sehen die Maus; dies aktiviert das Konzept MAUS. Sie müssen nun das zugehörige Lemma MAUS auswählen, jedoch *nicht* das Lemma SPINNE. Aber das Konzept MAUS schickt eine Aktivierung zum Konzept TIER und weiter zum Konzept SPINNE und seinem Lemma. Nun sind also beide Lemmata, MAUS und SPINNE, aktiviert. Tatsächlich werden so die Lemmata aller fünf Tiere aktiv. Sie konkurrieren miteinander, und das verursacht eine Verzögerung aller Tierantworten. Wenn aber das Bild der Maus in der heterogenen Testbedingung unter unähnlichen Bildern erscheint, zum Beispiel mit dem Zug (siehe Abbildung 9b), dann entsteht diese Aktivierungsausbreitung nicht, weil es keinen gemeinsamen, direkt übergeordneten Begriff gibt. Die Konkurrenz ist daher weniger scharf. Dies ist, was die Theorie, entgegen der Intuition der meisten Betrachter, vorhersagt: Langsamere Benennungen unter homogenen als unter heterogenen Bedingungen. Und genau dies haben wir gefunden (siehe Stabdiagramm zur Sprechlatenz in Abbildung 10).

Dies ist nur ein Beispiel unserer experimentellen Arbeit zur Auswahl von Wörtern beim Sprechen. Das Modell, das ich zusammen mit Ardi Roelofs, Antje Meyer und vielen unserer Studenten entwikkelt habe, beinhaltet eine genaue Darstellung des Zeitverlaufs der Wortwahl unter Konkurrenzbedingungen.

Kurze Einblicke in den Wettbewerb zwischen den Wörtern eröffnen uns auch gelegentliche Entgleisungen der Wortwahl. Selektionsfehler, wie wir sie oben gesehen haben, zeigen, daß ein bedeutungsverwandtes Wort (»blau« statt »grün«, »Gutes« statt »Schlechtes«, usw.) so stark aktiviert war, daß es das Zielwort ersetzte.

Es gibt auch seltene Fälle, in denen zwei konkurrierende Wörter gemeinsam ausgewählt und dann miteinander verschmolzen wer-

den. Oft handelt es sich bei diesen Verschmelzungen um beinahe synonyme Wörter.

Wortverschmelzungen

Wir wollen keine ellenlose Debatte (ellenlange / endlose).
Denn die Sache ist kein Sperz (Scherz / Spaß)
Was hält die Koalition noch zustande? (zustande bringen / zusammenhalten) (aus dem Bierwisch Corpus)

Die Einheiten, die wir bei der Auswahl aus dem Lexikon verwenden, sind nicht nur Wörter, sondern auch Redewendungen, wie *um den heißen Brei herumreden*, *den Kopf verlieren* oder *guter Laune sein*, und Sprichwörter wie *am Abend wird der Faule fleißig* oder *gut Ding will Weile haben*. Ungefähr 8 % aller Wörter, die wir aussprechen, gehören zu solchen Redewendungen. Wir haben im Experiment zeigen können, daß sich auch bei der Auswahl solcher Redewendungen Konkurrenz ergibt. Und genau wie bedeutungsähnliche Wörter verschmelzen manchmal auch ähnliche Redewendungen, wie in diesen Beispielen:

Verschmelzungen von Redewendungen

Zur Frage gekommen (zur Sprache gekommen / in Frage gestellt worden).
Er setzt sich auf den Hinterkopf (er setzt sich in den Kopf / er stellt sich auf die Hinterbeine). (aus dem Meringer corpus)

Damit befinden wir uns bereits mitten in der syntaktischen Planung und Anordnung, einem besonders spannenden Kapitel unserer Wissenschaft, das ich hier aber leider nicht weiter erörtern kann. Das Ergebnis dieser syntaktischen Prozesse ist die Bildung von Phrasen oder Satzteilen.

Lassen Sie uns jetzt die äußere Sprachtätigkeit unserer Sprecherin Evelyn betrachten. Evelyn erzeugte eine Abfolge von gesprochenen Silben. Silben sind die artikulatorischen Einheiten unserer Sprache. Wie schnell produzierte Evelyn ihre Silben? Sehen Sie sich dazu Abbildung 5c an.

Ich habe einen Strich bei jeder Silbe gesetzt. Das sind drei bis vier Silben pro Sekunde. Die Silben bestehen ihrerseits aus Konsonanten und Vokalen. Was ist nun die Geschwindigkeit dieser elementaren Sprachlaute? Das zeigt uns Abbildung 5d, in der jeweils ein Strich für jeden Sprachlaut steht. Die durchschnittliche Geschwindigkeit ist etwa zehn Sprachlaute pro Sekunde, aber Sprints bis zu 15 Lauten pro Sekunde sind ganz normal. Die Hände der allerbesten Pianisten erreichen etwa 15 Anschläge pro Sekunde. In unserer Sprache erreichen wir alle diese Geschwindigkeit, und zwar ohne jegliches intensives Sprachtraining. Wie bereitet sich ein Sprecher auf diesen Strom von Sprachlauten vor? Er kann ihn unmöglich wirklich *planen*, im Sinne von – »soll ich nun ›p‹ oder ›t‹ sagen«? Es geht hier um eine völlig automatisierte Fertigkeit. Wie ist sie organisiert?

Es handelt sich um zwei superschnelle Prozesse, die einander auf den Fersen folgen. In unserem Grundriß, Abbildung 4, sind sie als zwei braune Felder dargestellt. Ich beschreibe sie ihnen nacheinander. Achten Sie aber zuerst einmal besonders auf die Position der phonologischen / phonetischen Kodierung im Grundriß. Zwischen dem begrifflichen Organ der inneren Sprache (die blauen Felder) und den motorischen Organen der äußeren Sprache (die braunen Felder) liegt so etwas wie eine Kluft, sowohl in psychologischer Hinsicht als auch neurologisch, in der Sprachentwicklung und selbst in der Evolution.

Betrachten wir zunächst das *phonologische Kodieren*. Um die flüssige Aussprache einer Abfolge von Wörtern vorzubereiten, muß der Sprecher für jedes einzelne Wort die Form, das letzte (in der Abbildung braune) Element im Lexikoneintrag, abrufen. Manchmal ist das nicht ganz einfach. Etwas dramatisch ausgedrückt: der Sprecher

muß die Kluft im System überwinden, nämlich vom kognitiven zum motorischen Feld. Sie kennen alle die peinliche Situation, wenn uns plötzlich ein Wort nicht mehr einfällt, etwa der Name einer Person, einer Stadt oder einer Pflanze. Wenn Ihnen so das Wort auf der Zunge liegt, dann können Sie die Kluft nicht überwinden.

Aber unsere Theorie sagt vorher, daß wir in dieser unglücklichen Lage die syntaktischen Merkmale des gesuchten Wortes kennen, weil das Lemma, das zweite Element der Triade, noch richtig ausgewählt wurde. Und tatsächlich ist das oft der Fall. Als Sprecher des Deutschen kennen Sie zum Beispiel oft das grammatische Geschlecht des gesuchten Wortes. Sie wissen also, daß es sich zum Beispiel bei dem Pflanzennamen um ein *der-*, *die-* oder *das-*Wort handelt. Experimentell ist dies mit Sprechern des Italienischen nachgewiesen worden.

Besonders bei selten vorkommenden Wörtern kann es schwierig sein, die Kluft zu überwinden. Die Worthäufigkeit spielt eine wichtige Rolle. Beachten Sie Abbildung 11. Sie können das Bild einer Nase zum Beispiel viel schneller – nun, etwa 30 Millisekunden schneller – benennen als das eines Hasen. »Nase« kommt in der Sprache viel häufiger vor als »Hase«. Dies ist der klassische Worthäufigkeitseffekt. Wir konnten zeigen, daß er vor allem beim Zugriff auf die Form der Wörter entsteht. Die Form ist, wie gesagt, die Kombination von Sprachlauten, die im Lexikon gespeichert ist. Man kann den Zugriff zu dieser Form eines Wortes beim Bildbenennen erleichtern, indem man den Sprecher über Kopfhörer ein ähnlich klingendes Wort hören läßt, d. h. ein Wort mit ein, zwei oder mehr übereinstimmenden Sprachlauten. Sie können also das Bild: der »Hase«, schneller benennen, wenn Sie gleichzeitig »Hafen« hören, als wenn Sie ein unähnliches Wort, zum Beispiel »Pulli«, hören (siehe Abbildung 12). Je mehr übereinstimmende Sprachlaute es gibt, desto größer ist die Beschleunigung.

Das phonologische Kodieren besteht vor allem aus der Silbenbildung. Dies läßt sich am einfachsten im Französischen demonstrieren. Nehmen wir an, Sie planen die Phrase *avec une amie* (siehe Abbildung 13). Ein wichtiges Ergebnis aus einem langjährigen For-

schungsprogramm, das ich zusammen mit Antje Meyer and Ardi Roelofs durchgeführt habe, ist, daß die Silbenbildung *inkrementell*, stückchenweise, verläuft. Der Sprecher ruft zunächst die Formen der aufeinanderfolgenden Wörter als ganzes aus dem Lexikon ab. In der Abbildung sind die Formen für die drei Wörter *avec*, *une* und *amie* dargestellt. Anschliessend wird jedes Wort in seine Sprachlaute zerlegt. Bei der Silbenbildung werden dann die zerlegten Konsonanten und Vokale inkrementell wieder zusammengeführt. Dadurch werden nacheinander die Silben der Äußerung gebildet. Abbildung 13 zeigt die inkrementelle Bildung der Syllaben *a-ve-ky-na-mi*. Das wichtigste ist hier, daß die Wörter bei der Silbenbildung ihre Eigenständigkeit verlieren können. Das geschieht hier für alle drei Wörter. Die Silben *ky* und *na* überschneiden die beiden Wortgrenzen. Sie waren keine Silben in den Ausgangswörtern. Silben werden während des phonologischen Kodierens inkrementell, abhängig von der Umgebung gebildet. Anders gesagt: In unserer äußeren Sprache gibt es keine Wörter mehr, sondern nur noch Silben. In langsamer deutscher Sprache werden Wortgrenzen seltener von Silben überschnitten als im Französischen, aber bei schnellerem Sprechtempo verlieren auch deutsche Wörter während der Silbenbildung oft ihre Eigenständigkeit.
Gelegentlich, allerdings erstaunlich selten, geht bei der Silbenbildung etwas schief, und dann macht man Fehler wie diese:

Phonologische Fehler

Eine *S*orte von *T*acher (Torte von Sacher)
Der d*i*m*o*nierende Faktor (d*o*m*i*nierende)
(aus dem Meringer corpus)

Wenden wir uns jetzt der zweiten Komponente der äußeren Sprache, dem *phonetischen Kodieren*, zu (siehe unteres braunes Feld in Abbildung 4). Um die zuvor gebildete Silbenfolge auszusprechen, muß man nacheinander die Sprechbewegungen für jede einzelne Silbe

planen (*a-ve-ky-na-mi*). Wie geht das? Aus statistischen Analysen meines ehemaligen Studenten Niels Schiller wissen wir, daß ich etwa 85 % meines Sprechens mit nur 500 verschiedenen Silben bewerkstellige. (Dabei kennt das Deutsche mehr als 10.000 verschiedene Silben). Diese wenigen Silbenmuster haben wir im Laufe unseres Lebens millionen Male produziert. Deshalb ist es sehr wahrscheinlich, daß sie als komplexe Motorprogramme gespeichert sind. Den Speicher für Silbenprogramme, den ich hier annehme, habe ich das »mental syllabary«, Silbarium, genannt. Abbildung 4 zeigt ihn rechts unten im Grundriß. Solche Gedächtnissysteme zeigen in der Regel Häufigkeitseffekte: Einträge, die *oft* abgerufen werden, können auch relativ *leicht* abgerufen werden. Das haben wir bereits bei den Wortformen gesehen. Wenn es also einen Speicher für Silbenprogramme, ein Silbarium, gibt, dann sollte man vorhersagen, daß häufige Silben schneller abgerufen werden können als seltenere Silben. Dies nachzuweisen war ein experimenteller Alptraum, der nach der Veröffentlichung unseres ersten Artikels zum Silbarium noch ein ganzes Jahrzehnt gedauert hat. Aber schließlich ist es meiner früheren Studentin Joana Cholin und mir gelungen.

Wenn die erste Silbe einer Äußerung häufig ist, dann beginnt die Äußerung schneller, als wenn die erste Silbe seltener ist. In Abbildung 14 sehen Sie links den statistisch bedeutsamen Silbenhäufigkeitseffekt für einsilbige Äußerungen des Niederländischen, rechts die Effekte für zweisilbige Äußerungen. Beachten Sie, daß in zweisilbigen Äußerungen nur die erste Silbe einen Häufigkeitseffekt zeigt. Die Frequenz der zweiten Silbe hat keinen Einfluß auf die Sprechlatenz. Das zeigt, daß der Sprecher mit der Aussprache beginnen kann, sobald er die erste Silbe eines Wortes oder einer Äußerung aus dem Silbarium abgerufen hat. Dies trifft allerdings nicht auf alle Sprachen zu. Englisch zum Beispiel ist eine Ausnahme, wie Joana Cholin in einem vor kurzem veröffentlichten Artikel gezeigt hat. Offenbar hängt im Englischen die Artikulation der ersten Silbe oft vom Charakter der zweiten Silbe ab, und deshalb warten englische Sprecher in der Regel mit dem Beginn der Artikulation, bis sie beide Silben abgerufen haben.

Wenn wir nun die abgerufenen Silbenprogramme schnell nacheinander ausführen, dann erzeugen wir endlich flüssige Sprache. Dabei muß die Tätigkeit von mehr als hundert Muskeln koordiniert werden, so daß sie blitzschnell und höchst präzise zusammenarbeiten. Dieses sogenannte »Ansatzrohr«, unser artikulatorischer Apparat, ist ganz unten rechts im Grundriß (Abbildung 4) dargestellt.

Selbstkontrolle

Wir haben nun die vier Hauptkomponenten der Sprechplanung, die von Konzepten bis zur Aussprache führen, freigelegt. *Eine* wesentliche Komponente fehlt allerdings noch in unserem Grundriß des Sprechens. Betrachten Sie bitte mal, was hier beim Sprecher passiert:

Versprecher-Korrektur

Nun wissen Eingeweihte, daß Gerhard Schröder sich für Berlin eine Ampelkoalition wünscht, also eine, (.) *art-rote-rone-grote-grüne* (.) *ch-rot-rote-k-Regierung zusammen* (.) *nein* (.) *fangen wir noch mal neu an eine* rot-grüne Regierung zusammen mit den Liberalen

(Aufnahme Thomas Christes)

Wenn wir sprechen, sind wir unsere eigenen Hörer. So können wir unsere Sprache »abhören« und notfalls verbessern. Immer wenn wir sprechen, kontrollieren wir unsere Sprache auch mehr oder weniger gründlich. Diese »Monitorfunktion« steht in unserem Grundriß links als Pfeil angedeutet. Wir konnten zeigen, daß wir bei dieser Selbstkontrolle das normale Sprachwahrnehmungssystem einsetzen, also dasselbe System, mit dem wir auch die Äußerungen anderer Sprecher wahrnehmen und verstehen. Der Hörer in uns ist immer aktiv, wenn wir sprechen, und das ist nicht immer ein reines Vergnügen. Ich zitiere ein Ordensmitglied:

Wenn man den eigenen Worten
eine Zeitlang zuhört,
wie sie dröhnen im eigenen Kopf –
man möchte die Augen zudrücken
wie ein kleines Kind,
sich die Ohren zuhalten
und am liebsten gar nichts mehr sagen.
Aber das wäre falsch.

(Enzensberger)

Das wäre *wirklich* falsch. Glücklicherweise können wir uns mit diesem Zuhören korrigieren. Wir haben entdeckt, daß solche Korrekturen recht systematisch aufgebaut sind. Das ist erstaunlich, denn der Sprecher hat ein großes Problem und sucht einen Ausweg. Aber in einer Hinsicht sind beinahe alle spontanen Selbstkorrekturen wohlgeformt: Der Sprecher behält die zugrundeliegende grammatische Struktur der unterbrochenen Äußerung bei. Die letztendliche Korrektur wird grammatisch und phonologisch haargenau an die unterbrochene Äußerung angepaßt. Dadurch wird die Korrektur ein genauer, wohlgeformter Abschluß der ursprünglichen unterbrochenen Äußerung.
In meinem Vortrag habe ich das Phänomen wie folgt akustisch demonstriert. Wenn ich aus der obigen Versprecher-Korrektur das problematische Stück, hier in *kursiv,* herausschneide und den richtigen Anfang und das korrekte Ende zusammenfüge, dann sollte sich, wenn das eine Stück in der Tat ein syntaktisch korrekter Abschluß des anderen ist, ganz normale Sprache ergeben. Das Resultat:

> Nun wissen Eingeweihte, daß Gerhard Schröder sich für Berlin eine Ampelkoalition wünscht, also eine rot-grüne Regierung zusammen mit den Liberalen.

Hört sich tatsächlich wie ganz normale, fließende Sprache an. Auch die Intonation klingt natürlich. Das ist kein Zufall. Ausnahmen sind

selten. Ich kann Ihnen noch ein weiteres akustisches Beispiel gegeben. Hier folgende Versprecher-Korrektur:

Versprecher-Korrektur

Der Weinheimer Volker Weidler ist bei der *Vorqualifa-g-eh-bei. Entschuldigung* (.) *bei der Vorqualifak-eh- bei-bei der* Vorqualifikation zum Training für den Großen Preis von San Marino (.) ausgeschieden.

Wenn ich das problematische Stück (*kursiv*) aus der Audioaufnahme herausschneide:

> Der Weinheimer Volker Weidler ist bei der Vorqualifikation zum Training für den Großen Preis von San Marino (.) ausgeschieden.

Hört man wieder ganz normale Sprache. Selbst ein Sprecher in Not ist gut organisiert.
Bisher haben wir die vier mentalen Spieler betrachtet, die zusammen unsere Sprachtätigkeit hervorbringen (Abbildung 4). Zwei davon erzeugen unsere innere Sprache. Die anderen beiden stellen unsere äußere Sprache bereit, diese komplexen, superschnellen Sprechbewegungen. Diese vier Komponenten arbeiten relativ unabhängig voneinander; das heißt, es gibt kaum schnelle Rückkoppelungen zwischen ihnen. Insbesondere wird die Produktion innerer Sprache selten von der äußeren Sprachtätigkeit beeinflußt. Rückmeldung erhalten wir vor allem dadurch, daß wir unsere eigene Sprache abhören. Dieser Rückkoppelungsprozeß ist recht langsam und nicht besonders zuverlässig; er überhört viele unserer Fehler und Ungenauigkeiten.

Wenn wir sprechen, sind die vier Agenten immer gleichzeitig am Werk. Wir planen den weiteren Inhalt unserer Äußerung meist schon, während wir den vorangegangenen Teil grammatisch in Phrasen umsetzen. Wir bilden diese Phrasen, während wir die vorangehenden Phrasen in Silben umsetzen. Und wir bilden diese Silben, während wir die vorangehenden Silben aussprechen. Diese inkrementelle, dachziegelartige Arbeitsweise ist eine faszinierende Eigenschaft dieser geistigen Hochgeschwindigkeitsmaschine.
Kehren wir noch einmal zu Evelyn zurück, während sie gleichzeitig auf allen vier Ebenen arbeitet und dabei Ideen, Phrasen und Wörter, Silben und Sprachlaute erzeugt. Diese vier Ebenen (Abbildung 5a, 5b, 5c und 5d) sind alle gleichzeitig aktiv. In dem Vortrag habe ich Sie Evelyns laufende Sprache zugleich mit den Klicklauten auf den vier Eben: Propositionen, Wörter, Silben und Laute hören lassen. Beim Sprechen sind wir uns dieser unvorstellbaren simultanen Aktivität nie bewußt.

Dank

In diesem Vortrag habe ich Ihnen beschrieben, wie wir in unserer Forschung die erstaunliche Fertigkeit des Sprechens in ihre Bestandteile zerlegt haben. Ich hatte das große Vorrecht, dabei im Laufe der Jahre mit buchstäblich Dutzenden hervorragender Wissenschaftler und Wissenschaftlerinnen – Doktoranden, Mitarbeiter und Kollegen – zusammenzuarbeiten. Dieser Vortrag ist ihnen allen gewidmet, besonders Antje Meyer, Ardi Roelofs und Niels Schiller.

Ich danke Ihnen.

Abb. 1: Wilhelm von Humboldt (1767-1835), Joseph Karl Stieler via Wikimedia Commons

Abb. 2a: Wilhelm Wundt (1831-1920)

Abb. 2b: James McKeen Cattell (1860-1944), © Lafayette College Archives

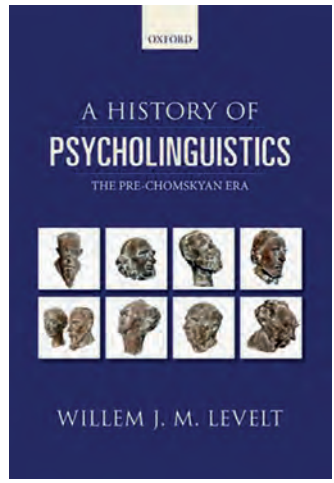
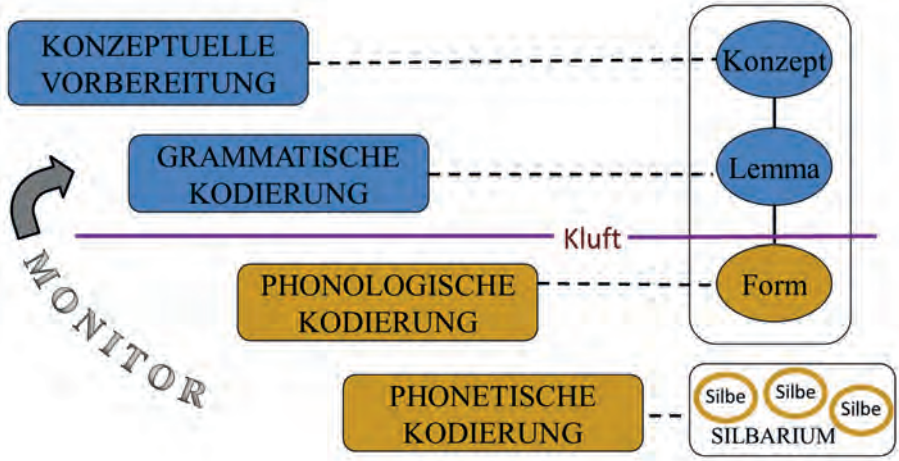


Abb. 3: A History of Psycholinguistics: The Pre-Chomsky Era. Oxford University Press, 2013



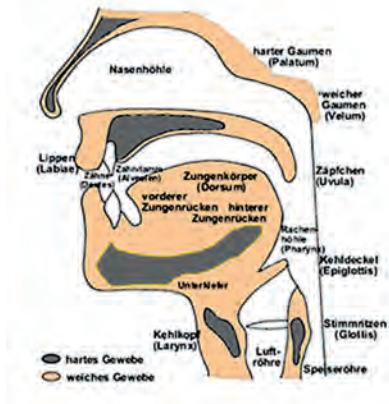


Abb. 4: Theoretischer Gesamtgrundriß des Sprachsystems, Querschnitt aus: Grießhaber, Wilhelm (o. J.) HyperPhonetik: Vokale, Konsonanten

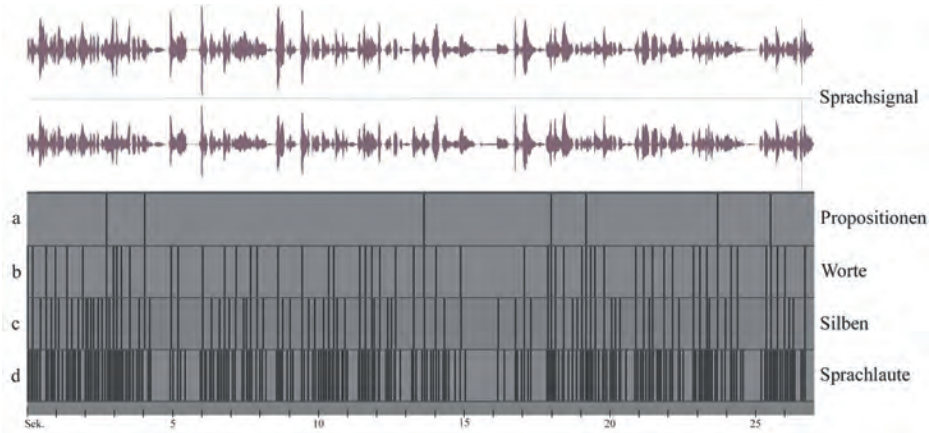


Abb. 5: Sprachsignal mit a. Tempo der Propositionen; b. Tempo der Wörter; c. Tempo der Silben; d. Tempo der Sprachlaute

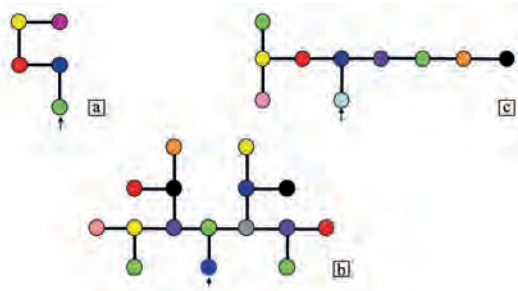


Abb. 6: Das Linearisierungsproblem. a. Erste Regel: Folge den Verbindungen!; b. Zweite Regel: Last-in, first-out; c. Dritte Regel: Das Einfachste zuerst!

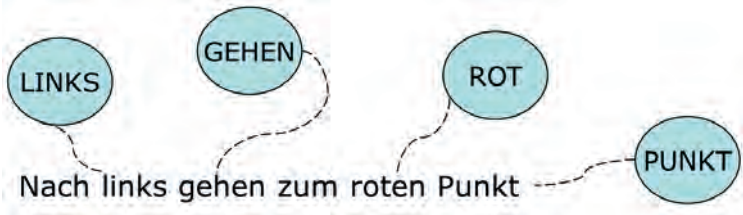


Abb. 7: Bereitstellung lexikalischer Konzepte

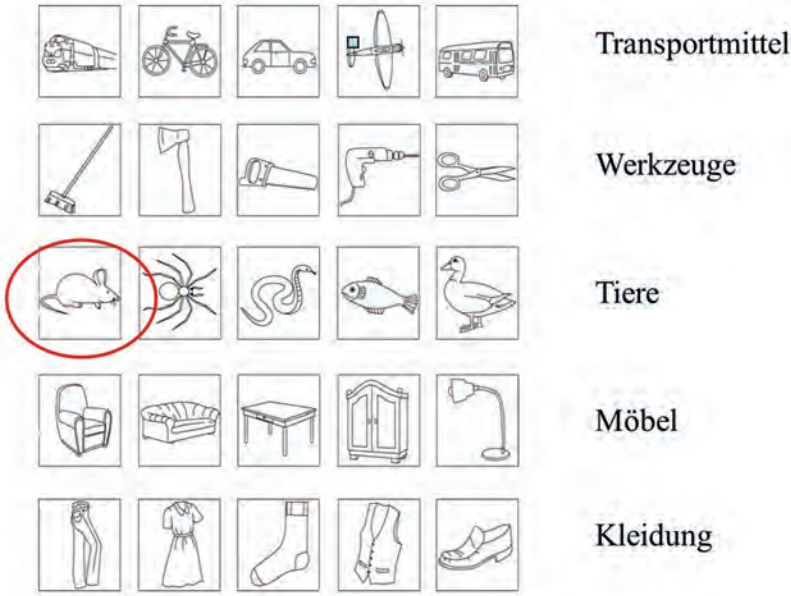


Abb. 8: Konkurrenz beim lexikalischen Zugriff

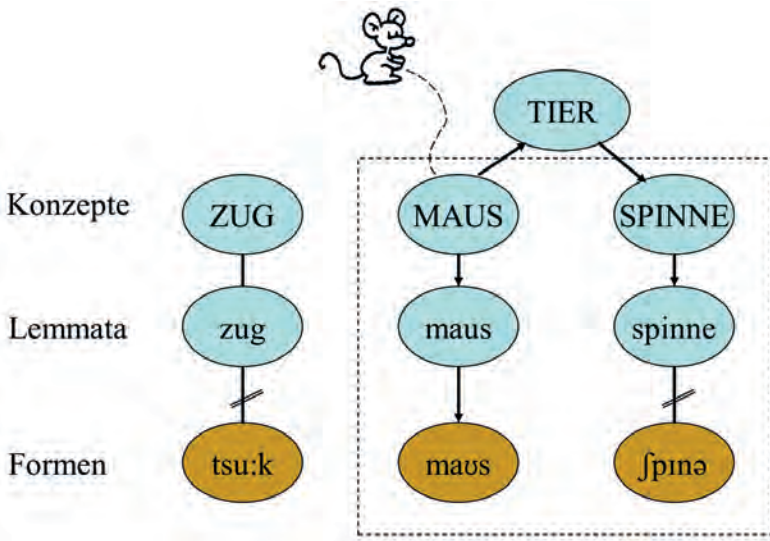


Abb. 9a: Konkurrenz beim lexikalischen Zugriff
unter homogenen Testbedingungen

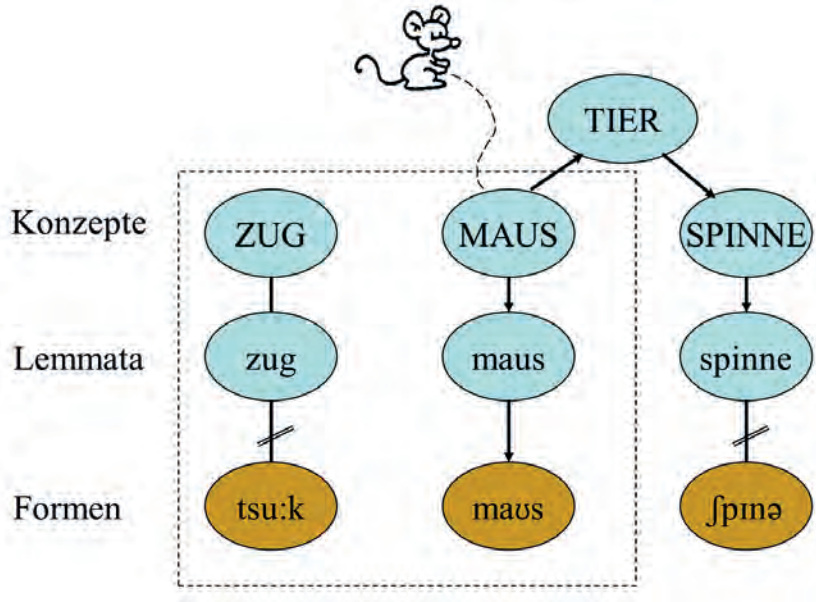


Abb. 9b: Konkurrenz beim lexikalischen Zugriff
unter heterogenen Testbedingungen

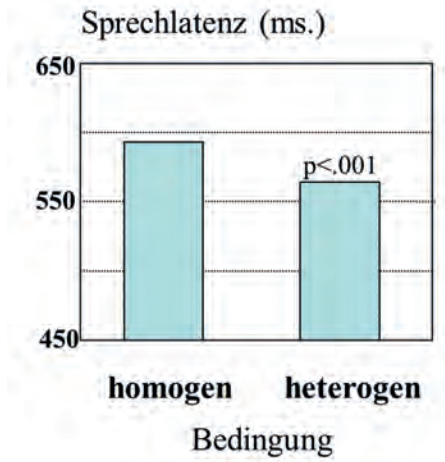


Abb. 10: Stabdiagramm zur Sprechlatenz unter homogenen und
heterogenen Bedingungen

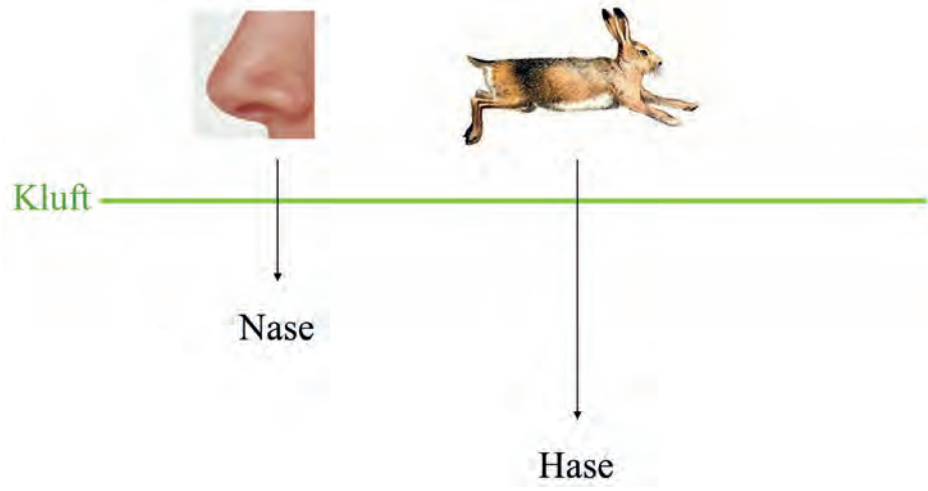


Abb. 11: Worthäufigkeitseffekt (Nase versus Hase)

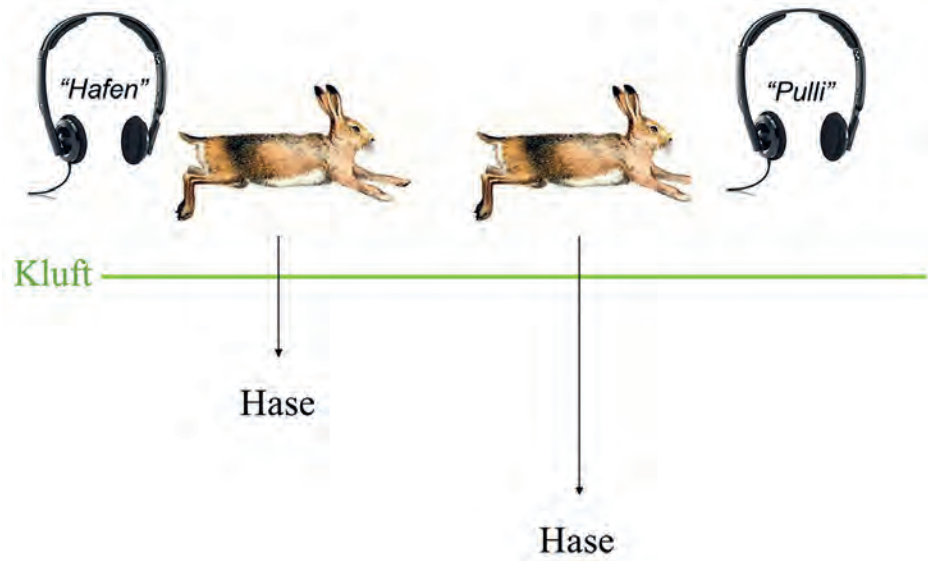


Abb. 12: Beschleunigung der Aussprache (von Hase)
beim gleichzeitigem Hören eines ähnlich klingendes Worts (Hafen)

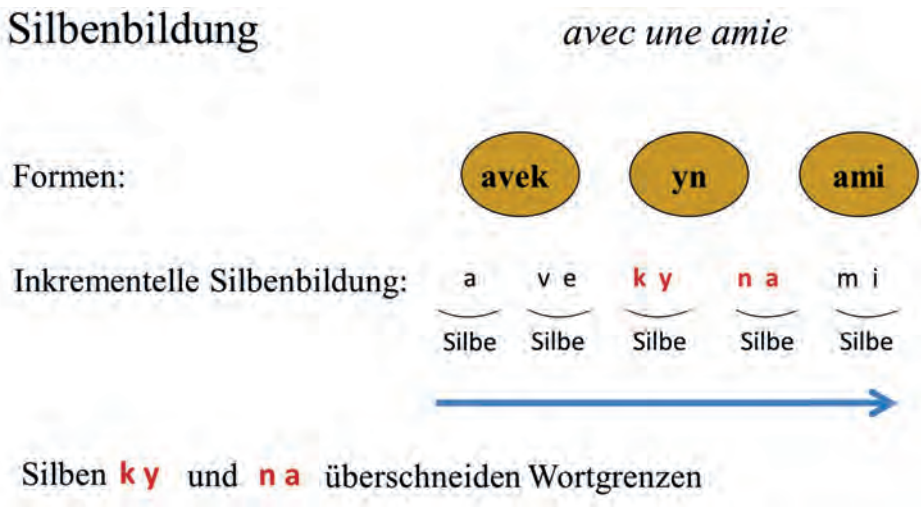


Abb. 13: Inkrementelle Bildung von Syllaben

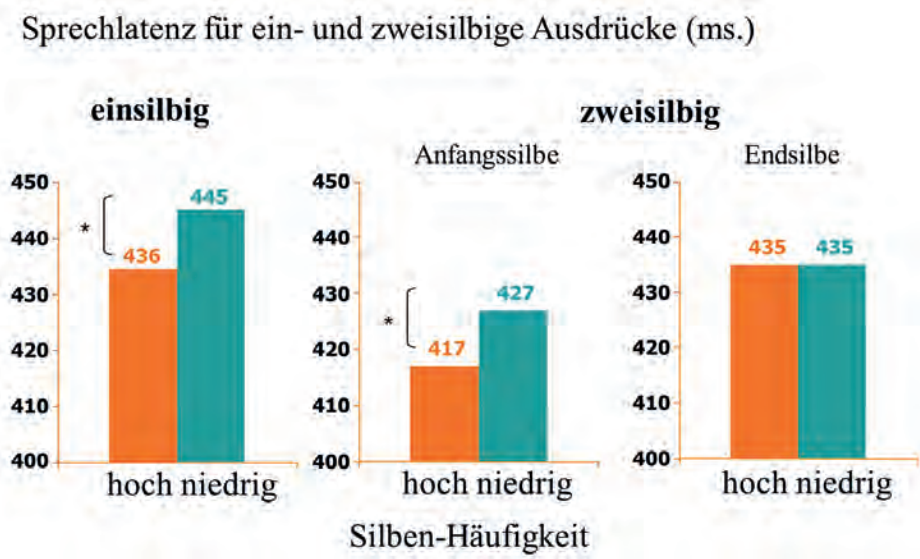


Abb. 14: Stabdiagramme zur Silbenhäufigkeit einsilbiger und zweisilbiger Ausdrücke, aus: Cholin, Levelt & Schiller (2006)