

Ein funktionalanalytisches Modell der Wortbildung

Andrea Krott

1. Einleitung

Die Funktionalanalyse als Erklärungsmethode ist darauf gerichtet, Eigenschaften von Systemen durch ihre Funktion im System zu erklären. Das im folgenden beschriebene Modell der Wortbildung, das auf der Funktionalanalyse basiert, untersucht daher die Funktion der Wortbildung im System Sprache. Dabei geht es zum einen um die allgemeine Funktion der Wortbildung, die sich als eine Methode zur Kodierung von Bedeutungen bestimmen läßt, also als ein Mittel zur Befriedigung des Kodierungsbedürfnisses, eines Teilaspektes des Kommunikationsbedürfnisses. Zum anderen geht es um die Analyse der *Systemgrößen* im Sprachsystem, welche mit der Wortbildung in Zusammenhang stehen. Dazu gehören beispielsweise das Morphinventar einer Sprache, die potentiellen Kombinationen eines Morphs mit anderen Morphen, die Anzahl der Neubildungen, die ein bestimmtes Morph enthalten, die Anzahl der Bedeutungen eines Morphs oder die Morphemlänge. Da es sich um eine funktionalanalytische Betrachtung handelt, stehen vor allem die Funktionen dieser Größen im System und damit die Einflüsse, welche sie aufeinander ausüben, im Mittelpunkt der Untersuchung. Aber auch die Einflüsse, die von außen auf das System einwirken, wie z.B. das Kodierungsbedürfnis, oder auch die Zusammenhänge der Größen mit anderen Teilen des Sprachsystems sind von Interesse.

Bei der folgenden Beschreibung des Modells werden die für die Wortbildung relevanten Systemgrößen der Sprache bestimmt und in Form von Hypothesen Abhängigkeiten zwischen ihnen formuliert.

Ein Teil der aufgestellten Hypothesen ist empirisch an den drei Sprachen Deutsch, Englisch und Niederländisch überprüft worden. Die Ergebnisse dieser Überprüfung werden im Anschluß an den deskriptiven Teil der Hypothesenbildung referiert. Auf diesen empirischen Teil wird besonderer Wert gelegt, da mit der vorliegenden Untersuchung ein Beitrag zum Aufbau eines umfassenden Modells der Struktur und Dynamik der Sprache geleistet werden soll. Dafür sind je-

doch umfangreiche empirische Untersuchungen notwendig, die weit über Beispielsätze bzw. Beispielwortbildungen hinausgehen.

2. Die Modellierung der Wortbildung

2.1 Systemgrößen und Systembedürfnisse in ihrem Beziehungsgefüge

2.1.1 Das Morphinventar

Das *Morphinventar* einer Sprache faßt alle kleinsten bedeutungstragenden Segmente auf der Ebene der *Parole* zusammen, die *Morphe*. Dabei werden für die folgenden Überlegungen und Untersuchungen die *Flexionsmorpheme* ausgeklammert, da sie in der Wortbildung allenfalls als *Fugenmorpheme* auftauchen, und es fraglich ist, ob man in solchen Fällen überhaupt noch von Flexionsmorphemen sprechen darf. Im Mittelpunkt der Untersuchungen zur Wortbildung sollen nicht Morpheme, sondern Morphe stehen, so daß Allomorphe getrennt behandelt werden und homonyme Morphe mit polysemen zusammenfallen. Dadurch wird das Problem vermieden, unter polysemen und homonymen Morphemen, die synchron nicht unterscheidbar sind, zu differenzieren.

Das Morph-Inventar ist aufgrund des dynamischen Charakters der Sprache einem ständigen Wandel ausgesetzt, der nicht zum Stillstand kommt, solange eine Sprache lebendig ist. Diese Lebendigkeit zeigt sich vor allem in der Kodierung neuer Bedeutungen, die auf das Kodierungsbedürfnis zurückzuführen ist. Zu den lexikalischen Methoden des Kodierungsbedürfnisses zählen die Erhöhung der *Wort-Polylexie* und die Lexikalisierung von Neologismen, Entlehnungen, Derivaten und Komposita bzw. anderen Wortbildungen (Köhler, 1990:105; Wimmer et al., 1994:105). Außer der Erhöhung der Wort-Polylexie, der Bedeutungsanzahl eines Wortes, vergrößern diese Methoden alle das Wortinventar einer Sprache. Die Aufnahme von Neologismen und Entlehnungen führt gleichzeitig auch zu einer Erweiterung des Morphinventars. Manchmal werden sogar Morpheme entlehnt, die nicht wortfähig sind: Um ein solches Fremdmorphem in der deutschen Sprache handelte es sich bei dem Lehnsuffix

-arius, das vermutlich im 8. Jahrhundert unter römischem Kultureinfluß übernommen worden ist (Erben, 1993:137). Derivation und Komposition führen normalerweise nicht zu einer Erweiterung des Morphinventars. Es kann jedoch vorkommen, daß wie im Beispiel *Milchgesicht* nur auf der Oberfläche Konstituenten erkennbar sind, daß jedoch diese Teile nichts mehr zu der Bedeutung des Wortes beitragen. Dann ist das Wort von Seiten seiner Bedeutung her nicht mehr in seine Konstituenten zerlegbar. Daher kann man festsetzen, daß es sich hier um ein neues Morphem handelt. Ähnlich sieht es bei einer Verschmelzung des Grundmorphems mit einem Ableitungsmorphem aus: Das *-l* in *Seil* war ursprünglich ein Suffix. In dem Wort *Seil* steckt die Wurzel *ši* (,binden‘) (Henzen, 1965:112f.). Durch die Verschmelzung ist ein neues Morphem entstanden.

2.1.2 Die Morph-Polysemie

Zu der Erhöhung der Morph-Polysemie zählt einmal die Bereicherung der Unterbedeutungen wie im Beispiel *Mäuse* = ‚Geld‘. Zum anderen wird auch der Wandel eines Grundmorphems zu einem Wortbildungsmorphem als Vermehrung der Bedeutungsanzahl des entsprechenden Morphs betrachtet. Auf Morphem-Ebene wird durch einen solchen Vorgang nicht die Bedeutungsmenge bereichert, sondern es entsteht ein neues Morphem. Auf der Morph-Ebene jedoch handelt es sich nach dem Wandel immer noch um dasselbe Morph, das allerdings eine neue Bedeutung bzw. Funktion erhalten hat.

Bei der Umsetzung eines Wortes in eine andere Wortart bleibt die ursprüngliche Bedeutung in der Regel erhalten wie in *Kalb* - *kalben*. Wenn sich das abgeleitete Wort jedoch verselbständigt, kann sich die Bedeutung mit der Zeit so sehr wandeln, daß die Ursprungsbedeutung nicht mehr erhalten ist. Ähnliche Phänomene lassen sich auch bei Kompositionen und Derivationen feststellen: Diese Art von Morphkombinationen können sich bei zunehmender Idiomatisierung von der Bedeutung der enthaltenen Morphe lösen, wobei zu erwähnen ist, daß in der Regel nicht alle Bedeutungen eines Morphs in der Kombination noch erhalten sind (vgl. Fleischer, 1976:59). Kann zwischen den ursprünglichen Bedeutungen der Teile und der Bedeutung der Kombination keine Beziehung mehr hergestellt werden, dann handelt es sich streng genommen um ein neues, selbständiges Morph; daher wird es auch in dieser Arbeit als eigenes Morph behandelt werden.

2.1.3 Der Morph-Inventarumfang

Bei der Kodierung einer Bedeutung auf Morph-Ebene, ohne auf das in der Sprache schon vorhandene Inventar zurückzugreifen, kann, wie schon erwähnt wurde, ein neues Morph aus einer anderen Sprache ‚entlehnt‘ oder in einer Neuschöpfung kreiert werden. Beides vergrößert den Morph-Inventarumfang. Hinzu kommt die Möglichkeit, ein Morph, wie im Falle mancher Kurz- und Initialwörter, zu kürzen. So soll davon ausgegangen werden, daß beispielsweise *Auto* und *Automobil* zwei verschiedene Morphe darstellen, da sie von ihrer Form her nicht identisch sind. Das Morphinventar vergrößert sich jedoch nicht nur, sondern Morphe können auch aussterben. Bei Henzen (1965:112, 113) finden wir beispielsweise den Hinweis, daß dies für ein Suffix gilt, wenn sein Bedeutungsgehalt so sehr verblaßt ist oder sein Lautkörper so sehr zusammenschrumpft, daß es das Ableitungsverhältnis nicht mehr genügend zu bezeichnen vermag oder Verwechslungen hervorrufen kann (Henzen, 1965:112, 113).

Ist eine Sprache durch eine durchschnittlich hohe Morph-Polysemie gekennzeichnet, so führt das bei gleichbleibenden Kodierungsanforderungen zu einem geringeren Morph-Inventarumfang als durch eine niedrigere durchschnittliche Morph-Polysemie. Aus diesen Zusammenhängen läßt sich folgende Annahme

über die Beziehung zwischen der durchschnittlichen Morph-Polysemie und dem Morph-Inventarumfang formulieren:

Hypothese 1: *Die Veränderungsrate des Morph-Inventarumfangs ist umgekehrt proportional zur durchschnittlichen Morph-Polysemie bei konstantem Kodierungsbedürfnis.*

Die Einflußstärke der Morph-Polysemie ist gekoppelt mit der Stärke des Bedürfnisses nach *Minimierung des Inventars (minI)*.

Die oben in Worten ausgedrückte Hypothese entspricht der Differentialgleichung

$$\frac{y'}{y} = \frac{-S}{x},$$

wobei y dem Morph-Inventarumfang und x der Morph-Polysemie entspricht. S ist ein Proportionalitätsfaktor und steht für die Stärke, die *minI* ausübt. Die Differentialgleichung besitzt die Lösung $y = C x^{-S}$ wobei C eine beliebige Konstante ist. Nach Logarithmierung der Gleichung und Ersetzung der Variablen x und y erhält man:

$$L\text{-MORPH-INV.-UMFANG} = c - S \cdot L\text{-ØMORPH-POLYSEMIE}.$$

Aus den obigen Überlegungen heraus läßt sich die Variable $c = \ln C$ als Einfluß des Kodierungsbedürfnisses interpretieren. Somit lautet die vollständige Gleichung:

$$L\text{-MORPH-INV.-UMFANG} = E \cdot \text{Kod} - S \cdot L\text{-ØMORPH-POLYSEMIE},$$

wobei E die Einflußstärke des Kodierungsbedürfnisses darstellt. Mit den Mitteln der Graphenalgebra läßt sich diese hypothetische Beziehung durch die nachstehende Abbildung ausdrücken.

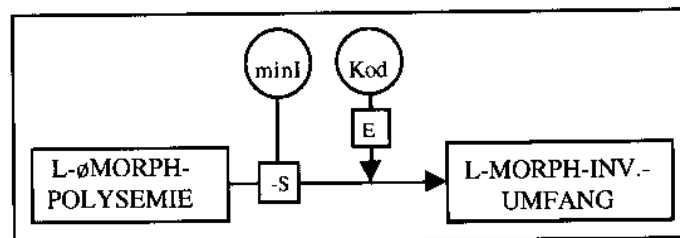


Abb. 2.1.3: Die Abhängigkeiten des Morph-Inventarumfangs

2.1.4 Das Potential der Morphkombinationen

2.1.4.1 Das uneingeschränkte Morph-Kombinationspotential

Bei der Wortbildung werden u.a. Morphe bzw. Morpheme miteinander kombiniert¹. Es lassen sich folgende Typen der Kombination unterscheiden: Komposition, Derivation, Iteration und Blending. Diese Wortbildungsmittel weisen in verschiedenen Sprachen unterschiedliche Grade der Ausnutzung auf. Beim Vergleich der deutschen mit der englischen Sprache stellt man beispielsweise fest, daß das Englische weniger Gebrauch von der Nominalkomposition macht als das Deutsche und auch eine geringere Neigung zu mehrgliedrigeren, motivierten Wörtern aufweist (Erben, 1993:129, 130). Dagegen findet man im Englischen viel häufiger Konversionen als im Deutschen (vgl. Fleischer, 1976:76)².

Die Kombinationen von Morphen unterliegen bestimmten Regeln oder Restriktionen. Davon soll jedoch zunächst einmal abstrahiert werden. Wörter werden in einem ersten Schritt als bloße Kombinationen von Morphen betrachtet. Die Gesamtheit aller Morphketten einer bestimmten Kombinationslänge soll *kombinatorisches Morph-Potential der Länge l* heißen. Wir betrachten also jeweils die möglichen Kombinationen einer festgelegten Anzahl von Morphen. Dieses Potential hängt von der Anzahl der Morphe ab. Das bedeutet, wenn mehr Morphe zur Kombination zur Verfügung stehen, können auch mehr Kombinationen gebildet werden.

Mit den Mitteln der Kombinatorik läßt sich das Morph-Kombinationspotential der Länge l (KOM_l) bei einem Morph-Inventar von n Morphen wie folgt berechnen: $KOM_l = n^l$. Eine Logarithmierung der Gleichung ergibt:

$$L-KOM_l = l \cdot \ln n \text{ bzw. } L-KOM_l = l \cdot L-MORPH-INV.-UMFANG.$$

Diese Beziehung ist mit graphischen Mitteln in Abbildung 2.1.4.1 dargestellt. Abbildung 2.1.4.1a faßt die bisher vorgestellten Systemgrößen und ihre Abhängigkeiten noch einmal zusammen.

¹ Um die Möglichkeiten der Wortbildung durch diejenigen Mittel zu vervollständigen, die nicht zu den Morphkombinationen gehören, müssen noch *Zeichenkürzung* und *implizite Ableitung* (die Umsetzung eines Wortes in eine andere Wortart) erwähnt werden. Unter *Zeichenkürzung* fallen *Clipping*, *Rückbildung* und *alphabetische Formation* wie *Buchstabenwörter* und *Akronyme* (vgl. Erben, 1993:57; Bartsch & Vennemann, 1983:166, 167).

² Auf eine ausführliche Beschreibung dieser Unterschiede soll in dieser Arbeit nicht eingegangen werden. Eine solche Untersuchung ist jedoch eine interessante Aufgabe, der eine eigene Arbeit gewidmet werden könnte.

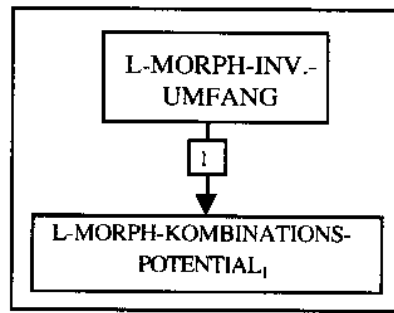


Abb. 2.1.4.1: Die Abhängigkeit des Morph-Kombinationspotentials der Länge l vom Morph-Inventarumfang

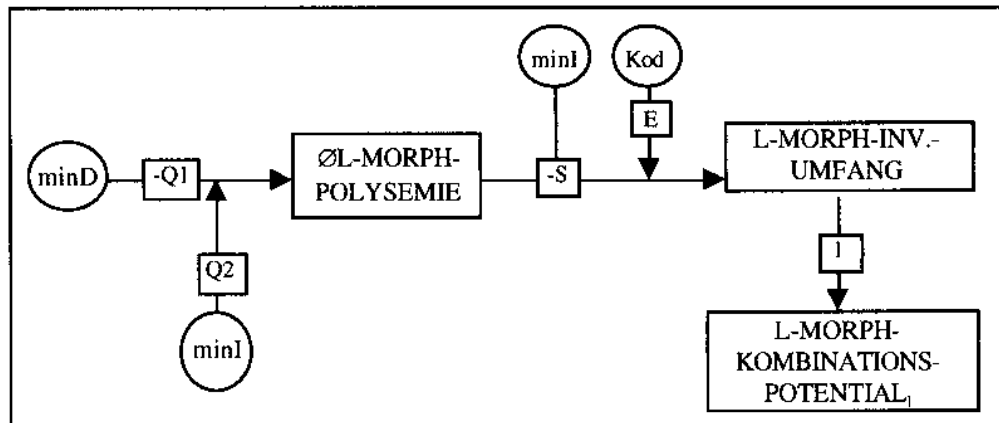


Abb. 2.1.4.1a: Die direkten und indirekten Einflüsse auf das Morph-Kombinationspotential der Länge l

Bei der Bestimmung des kombinatorischen Morph-Potentials der Länge l wurde noch nicht erwähnt, daß nicht alle möglichen Längen, die rein theoretisch auch unendlich sein könnten, in einer Sprache vorkommen. Es existiert vielmehr für jede Sprache eine eigene maximale Wortlänge. Die Summe der kombinatorischen Morph-Potentiale aller vorkommenden Wortlängen einer Sprache soll *kombinatorisches Morph-Potential (KOM)* einer Sprache heißen. Sie berechnet sich wie folgt:

$$KOM = n^{l_{\max}} + n^{l_{\max}-1} + n^{l_{\max}-2} + \dots + n^2 + n = \sum_{l=1}^{l_{\max}} KOM_l .$$

Es stellt sich jedoch die Frage, wie die maximale Wortlänge l_{\max} zustande kommt. Zur Erklärung dieser Längenbeschränkung lassen sich verschiedene Bedürfnisse anführen: Kürzere Wörter kommen sowohl dem Bedürfnis nach Minimierung des Dekodierungsaufwands (*minD*) als auch dem Bedürfnis nach Minimierung des Produktionsaufwands (*minP*) nach. Denn kürzere Wörter sind für den Hörer

leichter und schneller zu verstehen und stellen für den Sprecher einen geringeren Produktionsaufwand dar. Obwohl *minP* und *minD* in anderen Fällen entgegengesetzt wirken, stellen sie im Falle der Bevorzugung kurzer Wörter kooperierende Kräfte dar. Ein anderer Einfluß auf die maximale Wortlänge stellt vermutlich das Bedürfnis nach Minimierung des Gedächtnisaufwands dar, zumindest dann, wenn man davon ausgeht, daß kürzere Wörter mit geringerem Aufwand im Gehirn gespeichert werden können als lange.

2.1.4.2 Das restringierte Morph-Kombinationspotential

Unter Berücksichtigung einer Reihe von Restriktionen läßt sich aus dem kombinatorischen Morph-Potential das *restringierte Morph-Kombinationspotential* bestimmen, das für eine Sprache relevant ist. Zu den Restriktionen gehören erstens die *semantischen Restriktionen*. Eine derartige Einschränkung besteht z. B. bei der Kombination mit dem deutschen Suffix *-in*, das nur mit Basiswörtern verbunden wird, die eine Person oder ein Tier bezeichnen und das Merkmal „weiblich“ noch nicht besitzen. Aus semantischen Gründen ist laut Erben (1993:45) auch eine Bildung wie **zerblühen* ausgeschlossen. Derartige Beschränkungen existieren jedoch nicht nur bei Derivationen: Zu semantischen Restriktionen zählt auch, daß im Deutschen Synonyme nicht zu einem substantivischen Determinativkompositum zusammengefügt werden können. Daher ist eine Bildung wie **Liftfahrstuhl* auszuschließen (vgl. Stepanowa & Fleischer, 1985:81).

Eine andere Art von Restriktionen sind die *artikulatorischen Restriktionen*, die für die Vermeidung schwer auszusprechender Morph-Kombinationen stehen. Zu diesen Beschränkungen, die auf der Sprecherseite anzusiedeln sind, existiert eine Parallele auf der Hörerseite: Die *rezeptiven Restriktionen*. So würde die Kombination *trage-korb* der Morphkette *trag-korb* vorgezogen, da das im Auslaut verhärtete *g* von dem folgenden *k* bei einer 'durchschnittlichen' Sprechgeschwindigkeit für den Hörer schwer abzugrenzen wäre.

Es finden sich auch Restriktionen, die sich *nicht* den drei genannten Typen zuordnen lassen. Solche Restriktionen können durch die Wirkung des *Redundanzbedürfnisses* (*Red*) erklärt werden, das die Übertragungssicherheit garantiert. Es handelt sich einerseits um Stellungsrestriktionen (vgl. Stepanowa & Fleischer, 1985:80): Das Anhängen aller Ableitungsmorpheme an das Grundmorphem schränkt deren Kombinierbarkeit zwar ein, fördert jedoch gerade durch diese Tatsache ein schnelleres und sichereres Verstehen auf der Hörerseite. Bei der Kombination von Grundmorphemen im nominalen Bereich der Komposition existieren im Deutschen keine Stellungsrestriktionen. Gerade diese Tatsache macht die Analyse der Bedeutungsbeziehung zwischen den Teilen eines Kompositums so schwierig.

Zu der letztgenannten ‚Restklasse‘ von Restriktionen zählen auch Beschränkungen, die bei Derivationen auftreten, und die Naumann (1986:35, 36) *morphologisch* bzw. *morphophonologisch* nennt: Das deutsche Suffix *-ig* tritt bei

spielsweise nur in Verbindung mit heimischen Kernmorphemen auf. Der Unterschied zwischen *-heit* und *-keit* entspricht einer morphophonologischen Restriktion, da sie komplementäre Distribution aufweisen. Solche Einschränkungen nennt Baayen (1992) auch für das Niederländische: Er bemerkt, daß die Suffixe *-heid* und *-te* beide zur Substantivbildung gebraucht werden, daß *-heid* jedoch hauptsächlich mit komplexen Basiswörtern, *-te* nur mit monomorphemischen Basiswörtern verbunden wird. Auch im Englischen existieren derartige Restriktionen. So wird beispielsweise das Suffix *-ity* grundsätzlich mit Morphemen lateinischen Ursprungs verbunden (vgl. Aronoff, 1976:51). Zu den Einschränkungen zählt auch, daß Affixe fast immer wortklassenspezifisch sind: So dient das Suffix *-ar/-är* der Adjektivderivation und wird nur mit Substantiven verbunden, wie in *illusionär* oder *atomar*. Mit dem Suffix *-ung* werden Substantive, mit *-ieren* immer Verben gebildet.

Die Anführung von Restriktionen darf nicht darüber hinwegtäuschen, daß solche Einschränkungen nicht unumstößlich sind. Diachron betrachtet verändern sich diese Restriktionen. Dafür wollen wir uns im folgenden vor Augen führen, wie Restriktionen eigentlich entstehen und welche Rolle sie bei die Bildung eines neuen Wortes spielen.

Es wurde erwähnt, daß die Restriktionen das restringierte Morph-Kombinationspotential, also die Anzahl der möglichen Wörter einer Sprache, einschränken. Man kann daher eine konkrete Wortbildung beschreiben als die Wahl eines Wortes aus demjenigen Teil des restringierten Morph-Kombinationspotentials, der noch nicht lexikalisiert ist. Diese Modellvorstellung ist jedoch sehr wirklichkeitsfern. Kein Sprecher wird eine Menge von möglichen Wörtern im Kopf haben und daraus eines auswählen. Statt dessen sollte man sich die Bildung eines neuen Wortes als Analogiebildung vorstellen, die auf der Grundlage schon bekannter Wörter zu neuen Bildungen führt (vgl. auch Becker, 1990:31; Derwing, 1990:253; Bybee, 1985:114). Die Restriktionen sind eigentlich nur abstrakte Beschreibungen derjenigen Morphkombinationen, die in einer Sprache verwendet werden. Geht man bei der Wortbildung von Analogiebildungen aus, so können auch die folgenden, von van Marle (1990:268) angeführten Phänomene erklärt werden: Zufälliges Auslösen von existierenden derivativen Mustern, die nicht produktiv sind, Erzeugung neuer derivativer Muster und nicht-derivative Prägungen.

Verstöße gegen bestehende Kombinationsrestriktionen werden oft auch als stilistisches Mittel benutzt, beispielsweise in der Literatur. Zu solchen literatursprachlichen Abweichungen zählt die unübliche Suffixkombination *Fremdling+in*, wie sie bei dem Autor Trakl vorkommt (vgl. Erben, 1993:54).

Alle bestehenden Restriktionen wirken negativ auf das restringierte Morph-Kombinationspotential. Das heißt, je mehr Restriktionen vorhanden sind, um so weniger Morph-Kombinationen sind in einer Sprache möglich. Die Hypothese 2 drückt diesen Sachverhalt etwas genauer aus.

Hypothese 2: Die Veränderungsrate des restringierten Morph-Kombinationspotentials ist umgekehrt proportional zu den vorhandenen Restriktionen.

Diese Annahme entspricht der Differentialgleichung $\frac{y'}{y} = \frac{-U}{x}$ mit der Lösung $y = C \cdot x^{-U}$.

Hierbei steht x für die Restriktionen, y für das restringierte Morph-Kombinationspotential, und U stellt den Proportionalitätsfaktor dar. Ersetzt man das restringierte Morph-Kombinationspotential durch das Kürzel *RESTR* und logarithmiert die Gleichung, dann erhält man die folgende Beziehung:

$$L\text{-RESTR} = \ln C - U \cdot L\text{-RESTRIKTIONEN} .$$

Die Formel lautet bei einer ausführlichen Nennung der Restriktionen (*SEM* steht für die semantischen Restriktionen, *ART* für die artikulatorischen und *REZ* für die rezeptiven):

$$(i) \quad L\text{-RESTR} = \ln C - U_1 \cdot L\text{-SEM} - U_2 \cdot L\text{-ART} - U_3 \cdot L\text{-REZ} - U_4 \cdot RED .$$

Beim restringierten Morph-Kombinationspotential handelt es sich um eine Teilmenge der uneingeschränkten Morph-Kombinationspotentiale aller Kombinationslängen für eine Sprache. Daher kann deren Beziehung in der Gleichung

$$RESTR = \frac{1}{C} \cdot KOM$$

ausgedrückt werden, in der *KOM* für das uneingeschränkte Morph-Kombinationspotential und *RESTR* für das restringierte Morph-Kombinationspotential steht. Die Variable C ergibt sich aus der Einwirkung der kombinatorischen Restriktionen und des Redundanzbedürfnisses: Je mehr Restriktionen bestehen, um so größer ist der Wert von C , und um so weniger Elemente enthält das restringierte Morph-Kombinationspotential.

Nach Linearisierung mittels Logarithmierung lautet die letztgenannte Gleichung:

$$(ii) \quad L\text{-RESTR} = L\text{-KOM} - \ln C .$$

Die Berücksichtigung der beiden Gleichungen (i) und (ii) führt zu der endgültigen Formel:

$$L\text{-RESTR} = L\text{-KOM} - U_1 \cdot L\text{-SEM} - U_2 \cdot L\text{-ART} - U_3 \cdot L\text{-REZ} - U_4 \cdot RED .$$

Die Abbildung 2.1.4.2 bietet eine graphische Darstellung dieser Zusammenhänge.

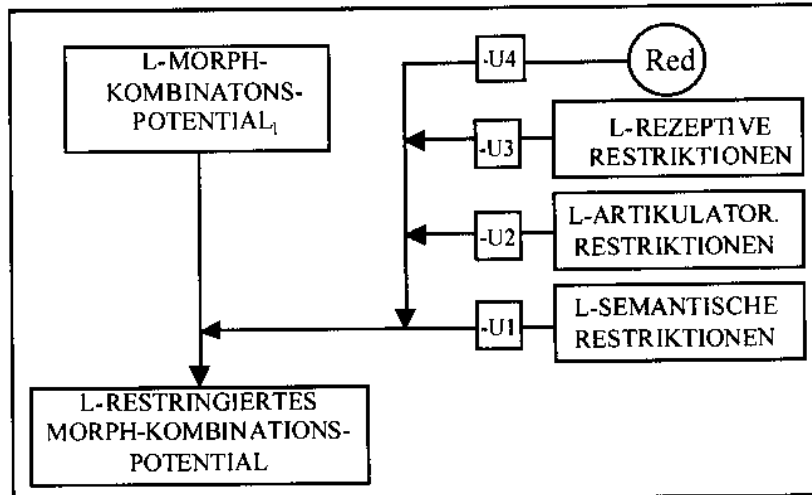


Abb. 2.1.4.2: Die Einflüsse auf das restringierte Morph-Kombinationspotential einer Sprache

2.1.4.3 Das lexikalisierte Morph-Kombinationspotential

Nachdem auf das Zustandekommen des restringierten Morph-Kombinationspotentials eingegangen worden ist, stellt sich die Frage, wie dieses Potential und die tatsächlich im Lexikon einer Sprache aufgenommenen Kombinationen zusammenhängen. Dabei betrachten wir nicht das gesamte restringierte Potential, sondern die Potentiale der einzelnen Morphe. Das restringierte Potential eines Morphs m soll *restringiertes Kombinationspotential des Morphs m* heißen. Die Lexikalisierung einer Morph-Kombination (eines Wortes) hängt sicherlich von dem Kodierungsbedürfnis, dem Bedürfnis, Bedeutungen auszudrücken, ab. Je mehr Bedeutungen eines Morphs für eine Sprachgemeinschaft relevant sind, um so größer ist das lexikalisierte Morph-Kombinationspotential. Diesem Prozeß wirkt jedoch das Bedürfnis nach Minimierung des Inventars (*minI*) entgegen, das mit dem Kodierungsbedürfnis in einem Fließgleichgewicht steht. Dieses Bedürfnis spiegelt die Erkenntnis wieder, daß die Kapazität des Gehirns begrenzt ist, und daher ein beschränktes Wortinventar von Vorteil ist.

Das *lexikalisierte Morph-Kombinationspotential des Morphs m* , d.h. die Menge der lexikalisierten Wörter, die das Morph m enthalten, bildet eine Teilmenge des restringierten Morph-Kombinationspotentials des Morphs. Dieser Zusammenhang wird durch die nachstehende Gleichung beschrieben:

$$LEX_m = \frac{1}{C} \cdot RESTR_m.$$

Dabei steht LEX_m für das lexikalisierte Morph-Kombinationspotential des Morphs m und $RESTR_m$ für sein restringiertes Kombinationspotential. Eine Logarithmierung der Gleichung ergibt

$$(iii) \quad L-LEX_m = L-RESTR_m - \ln C.$$

Die Komponente $\ln C$ setzt sich aus drei Teilkräften zusammen: Zunächst beinhaltet sie das erwähnte Fließgleichgewicht, das zwischen dem Bedürfnis nach Minimierung des Inventars und dem Kodierungsbedürfnis besteht. Hinzu kommt eine dritte Größe: Die Morph-Polysemie. Besitzt Morph A beispielsweise zehn Bedeutungen und Morph B eine Bedeutung, so ist zu vermuten, daß Morph A in mehr Wörtern des Lexikons aufgenommen ist als Morph B. Aus diesen Überlegungen heraus läßt sich folgende Hypothese formulieren:

Hypothese 3: Die Veränderungsrate des lexikalisierten Kombinationspotentials eines Morphs m ist proportional zur Morph-Polysemie des Morphs.

Diese Hypothese ähnelt einer Annahme von Altmann: „The greater the polylexy of a word the more compounds there are of which it is a component“³

Allerdings sagt Altmann etwas über Wörter und deren Vorkommen in Komposita aus. Da in dieser Arbeit jedoch nicht nur Komposita, sondern alle Arten der Wortbildung von Interesse sind, wurde versucht, mit der Zerlegung von Wörtern in Morphe eine allgemeinere Basis zu finden. Die Hypothese 3 entspricht der Differentialgleichung

$$\frac{y'}{y} = \frac{G}{x},$$

wobei y dem lexikalisierten Morph-Kombinationspotential und x der Morph-Polysemie entspricht, und G den Proportionalitätsfaktor darstellt. Die Lösung der Differentialgleichung lautet:

$$y = C \cdot x^G.$$

Logarithmiert man die Gleichung und ersetzt die Variablen x und y , so führt das zu der Formel

$$(iv) \quad L-LEX_m = c + G \cdot L-MORPH-POLYSEMIE_m \quad \text{mit } c = \ln C.$$

³ Diese Annahme wurde von Rothe (1988) an einer Stichprobe, die auf der Grundlage des Deutschen Wörterbuchs von Wahrig erhoben wurde, untersucht und bestätigt.

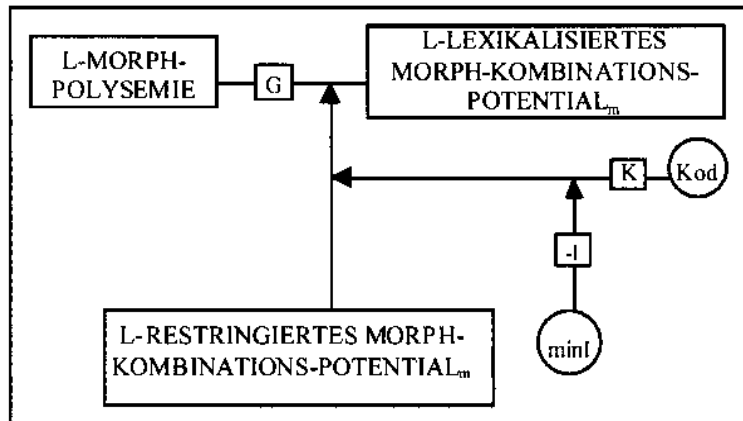


Abb. 2.1.4.3: Die Abhängigkeiten des lexikalisierten Kombinationspotentials des Morphs m

Faßt man die Gleichungen (iii) und (iv) zusammen, so läßt sich unter Berücksichtigung des Kodierungsbedürfnisses und des Bedürfnisses nach Minimierung des Inventars die endgültige Funktionsgleichung erstellen (MP_m steht für *Morph-POLYSEMIE_m*):

$$L-LEX_m = L-RESTR_m + K \cdot Kod - l \cdot \min l + G \cdot L-MP_m.$$

Diese Beziehungen werden auch in der Abbildung 2.1.4.3 dargestellt.

2.1.5 Die Morph-Frequenz

2.1.5.1 Die Morph-LexType-Frequenz

Das lexikalisierte Kombinationspotential eines speziellen Morphs übt einen Einfluß auf die Anzahl der in einem Korpus vorkommenden Wörter aus, die lexikalisiert sind und dieses Morph enthalten. Denn die Wahrscheinlichkeit, daß ein solches Wort im Text erscheint, steigt, je mehr Komposita oder Derivate dieses Morphs lexikalisiert sind.

Auf die Wörter im Korpus, die ein bestimmtes Morph enthalten, soll im folgenden mittels ‚Types‘ referiert werden. Die Anzahl der lexikalisierten Wörter, die ein bestimmtes Morph m enthalten und in einem Korpus verwendet werden, soll *Morph-LexType-Frequenz_m* heißen. Diese Type-Frequenz lexikalisierten Wörter hängt außer von dem genannten lexikalisierten Morph-Kombinationspotential auch vom *Anwendungsbedürfnis (Anw)* ab, d.h. von dem Bedürfnis, ein bekanntes Wort zu verwenden, welches das entsprechende Morph enthält. Dieses Anwendungsbedürfnis wird eingeschränkt durch das Vorhandensein synonyme Wörter. Je mehr solche Wörter bestehen, um so geringer ist die Wahrscheinlichkeit, eine Kombination des betreffenden Morphs zu verwenden.

Es wird also folgende Hypothese aufgestellt:

Hypothese 4: Die Veränderungsrate der Morph-LexType-Frequenz_m ist umgekehrt proportional zur Anzahl der synonymen Wörter.

In der Form einer Differentialgleichung lautet diese Annahme

$$\frac{y'}{y} = \frac{-L}{x}$$

Diese Gleichung besitzt die Lösung $y = C \cdot x^{-L}$. Variablenersetzung und Logarithmierung führt zu der Funktion:

$$(v) \quad L\text{-LEXTYPE-FREQU}_m = \ln C - L \cdot L\text{-SYNONYMIE}.$$

Da die Morph-LexType-Frequenz eine Untermenge des lexikalisierten Morph-Kombinationspotentials darstellt, läßt sich folgende Gleichung formulieren:

$$LEXTYPE - FREQU_m = \frac{1}{C} \cdot LEX_m,$$

wobei $LEXTYPE - FREQU_m$ für die Morph-LexType-Frequenz_m und LEX_m für das lexikalisierte Kombinationspotential des Morphs m steht. Logarithmiert man auch diese Formel, so lautet die Gleichung:

$$(vi) \quad L\text{-LEXTYPE-FREQU}_m = L\text{-LEX}_m - \ln C.$$

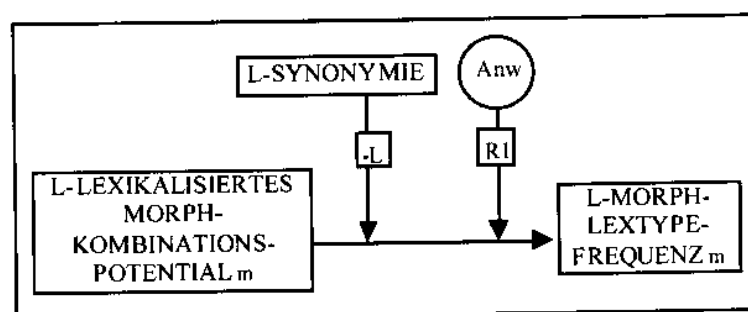


Abb. 2.1.5.1: Die Abhängigkeiten der Morph-LexType-Frequenz des Morphs m

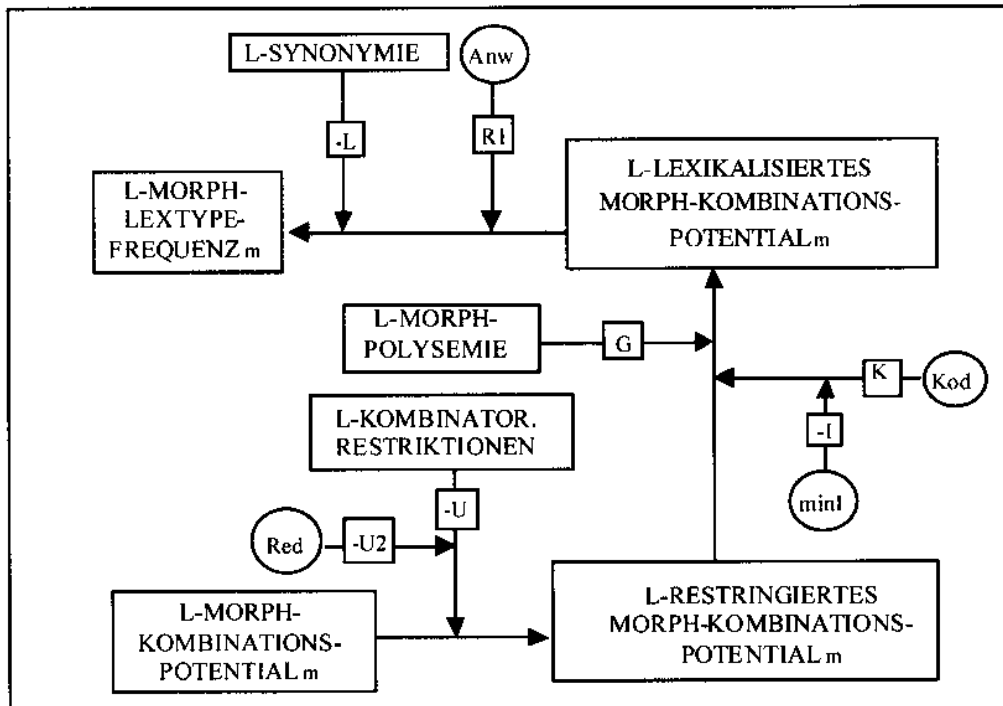


Abb. 2.1.5.1a: Die indirekte Abhängigkeit der Morph-LexType-Frequenz eines Morphs m von seinem Kombinationspotential

Die Konstante $\ln C$ läßt sich interpretieren als eine Verknüpfung der Einflüsse von Anwendungsbedürfnis und - unter Berücksichtigung der Gleichung (v) - der Synonymie:

$$L\text{-LEXTYPE-FREQU}_m = L\text{-LEX}_m + R_1 \cdot Anw - L \cdot L\text{-SYNONYMIE}.$$

Diese Formel entspricht dem in Abbildung 2.1.5.1 dargestellten Zusammenhang. Um die Einbettung der Morph-LexType-Frequenz in das System zu verdeutlichen, bietet Abbildung 2.1.5.1a einen Überblick über die indirekte Abhängigkeit der Morph-LexType-Frequenz eines Morphs m von seinem Kombinationspotential.

2.1.5.2 Die Morph-NeuType-Frequenz

Bekanntlich kommen in einem Text nicht nur lexikalisierte Wörter vor, sondern auch Neubildungen. Um auf die Möglichkeit zu referieren, neue Wörter mit Hilfe eines Wortbildungsmittels zu bilden, wird in der Wortbildungsliteratur der Begriff *Produktivität* verwendet. Dieser wird auf verschiedenen Ebenen des Sprachsystems benutzt: Für Konjugations- und Deklinationstypen, Verbalklassen, syntaktische Konstruktionen, Satzmodelle, Wortbildungsverfahren oder -typen. In allen diesen Fällen meint Produktivität die Anwendung einer Methode, mit deren Hilfe Bedeutungen ausgedrückt werden. Für diese Arbeit ist die Produktivität der

Wortbildungsmittel interessant. Ein Wortbildungsmittel ist dann produktiv, wenn mit ihm neue Wörter gebildet werden, ansonsten ist es unproduktiv.⁴ Von Produktivität soll jedoch auch bei Grundmorphemen gesprochen werden: Ein Grundmorphem ist produktiv, wenn es als Bestandteil eines neuen Wortes erscheint. Dadurch wird der Unterschied zwischen Grund- und Wortbildungsmorphemen in dieser Hinsicht aufgehoben.

Bei einem Vergleich verschiedener Wortbildungsmittel oder Grundmorpheme hinsichtlich ihrer Produktivität ist der jeweilige Anteil der Neubildungen an der Gesamtheit der neuen Wörter von Interesse. Es können jedoch auch einfach die absoluten Werte, die Anzahl der Neubildungen, miteinander verglichen werden. Wenn zwei Mittel oder Morpheme dieselbe Bedeutung haben und sich distributionell nicht unterscheiden, sind sie funktional äquivalent. Dann ist es möglich, einen direkten Vergleich der Produktivität vorzunehmen, der etwas über die Bevorzugung des einen Mittels oder Morphs gegenüber dem anderen aussagt.

Es stellt sich nun die Frage, warum ein Morph oft in einem neuen Wort erscheint und ein anderes nur selten oder gar nicht. In dem vorgestellten Zusammenhang heißt das: Welche Systemgrößen und -bedürfnisse wirken auf die Anzahl der Neubildungen in einem Text, für die ein bestimmtes Wortbildungsmittel bzw. Grundmorphem herangezogen wurde?

Als erstes ist hier die Zahl der potentiellen Neubildungen (POT_m) zu nennen, die sich aus der Differenz von restringiertem Morph-Potential des Morphs m ($RESTR_m$) und seinem lexikalisierten Kombinationspotential (LEX_m) ergibt:

$$POT_m = RESTR_m - LEX_m.$$

Die tatsächlich gebildeten Wörter bilden zu der Menge dieser potentiellen Neubildungen eine Untermenge. Dieses Verhältnis drückt die Gleichung

$$NEUTYPE - FREQU_m = \frac{1}{C} \cdot POT_m$$

aus, in der $NEUTYPE - FREQU_m$ für die Anzahl der neugebildeten Wörter in einem Korpus steht, die das Morph m enthalten. Die Konstante C steht stellvertretend für zwei weitere Einflüsse auf die möglichen Bildungen. Eines davon ist das Kodierungsbedürfnis, das Bedürfnis, neue Bedeutungen zu kodieren. Ein anderer Einfluß auf die Anzahl der neugebildeten Wörter ist die Anzahl synonyme Morpheme. Je mehr solche Morpheme bestehen, um so geringer ist die Wahrscheinlichkeit, daß das untersuchte Morph verwendet wird. Es wird daher folgende Hypothese aufgestellt:

⁴ Es wird hier nicht von der oft verwendeten Dreiteilung produktiv, aktiv und unproduktiv ausgegangen, da diese Einteilung nicht klar definiert ist.

Hypothese 5: Die Veränderungsrate der Morph-NeuType-Frequenz_m ist umgekehrt proportional zur Anzahl der synonymen Morphe.

Diese Annahme läßt sich auch durch die Gleichungen $\frac{y'}{y} = \frac{-T}{x}$ mit der Lösung $y = C \cdot x^{-T}$ ausdrücken, wobei y der Type-Frequenz des Morphs und x der Anzahl der synonymen Morphe entspricht. Eine Linearisierung durch Logarithmierung und die Ersetzen der Variablen ergeben:

$$L\text{-NEUTYPE-FREQU}_m = \ln C - T \cdot L\text{-MORPH-SYNONYMIE}_m$$

Berücksichtigt man die Zahl der potentiellen Neubildungen (POT_m) und das Kodierungsbedürfnis, dann lautet die endgültige Gleichung:

$$L\text{-NEUTYPE-FREQU}_m = POT_m + V \cdot Kod - T \cdot L\text{-MORPH-SYNONYMIE}_m$$

Diese Zusammenhänge werden graphisch noch einmal in der Abbildung 2.1.5.2 dargestellt.

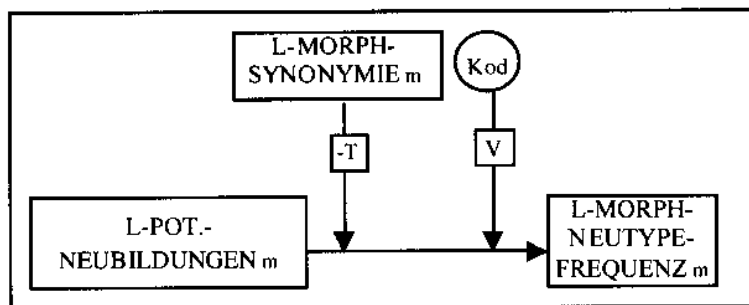


Abb. 2.1.5.2: Abhängigkeiten der Morph-NeuType-Frequenz_m.

2.1.5.3 Die Morph-Token-Frequenz

Die Summe der beiden Type-Frequenzen *Morph-LexType-Frequenz_m* und *Morph-NeuType-Frequenz_m* soll *Morph-Type-Frequenz_m* heißen. Diese gibt für ein einzelnes Morph an, in wievielen Types eines Korpus es enthalten ist. Von ihr wird angenommen, daß sie auf die Anzahl der zugehörigen Token, die *Morph-Token-Frequenz_m*, einwirkt. Das bedeutet, daß eine größere Anzahl von Types zu einem Morph vermutlich auch eine erhöhte Anzahl von Token zur Folge hat. Die erneute Verwendung eines Types wird jedoch durch das Bedürfnis nach Variabilität des Ausdrucks (Var) eingeschränkt. Infolgedessen wirkt das Vorhandensein synonymen Wörter vermutlich negativ auf die Morph-Token-Frequenz.

Das gilt beispielsweise für die konkurrierenden Formen *gewinkt* - *gewunken*⁵. Es existiert jedoch auch hierzu ein konkurrierendes Bedürfnis, nämlich das nach Vereinheitlichung der Benennung, das *Invarianzbedürfnis* (*Inv*)⁶. Dieses und ein weiterer Einfluß, das Anwendungsbedürfnis, wirken positiv auf die Anzahl der Token eines Wortes. Nimmt man alle diese Einflüsse zusammen, dann wird verständlich, warum ein Morph wie *und*, das nur sehr selten Kombinationen eingeht, wie in *tausend+und+ein*, trotzdem eine hohe Token-Frequenz aufweist. Denn das Anwendungsbedürfnis für *und* ist sehr hoch und die Anzahl der synonymen Ausdrücke sehr gering (z. B. *sowohl ... als auch*).

Die Zusammenhänge zwischen Type-Frequenz bzw. Synonymie und Token-Frequenz sind in den folgenden Hypothesen 6 und 7 noch einmal ausgedrückt:

Hypothese 6: *Die Veränderungsrate der Token-Frequenz eines Morphs m ist proportional zu seiner Type-Frequenz.*

Die zugehörige Differentialgleichung lautet

$$\frac{y'}{y} = \frac{M}{x} \text{ und besitzt die Lösung } y = C \cdot x^M \text{ bzw.}$$

$$(vii) \text{ TOKEN-FREQU}_m = C \cdot \text{TYPE-FREQU}_m^M$$

Hypothese 7: *Die Veränderungsrate der Token-Frequenz eines Morphs m ist umgekehrt proportional zur Synonymen-Anzahl der Morph-Types.*

In Form einer Differentialgleichung lautet diese Annahme:

$$\frac{y'}{y} = \frac{-H}{x} \text{ mit der Lösung } y = C \cdot x^{-H}$$

oder

$$\text{TOKEN-FREQU}_m = C \cdot \text{SYNONYMIE}^{-H}$$

⁵ In einem solchen Fall wäre es auch interessant, die Häufigkeiten der konkurrierenden Formen zu vergleichen, um etwas über die Bevorzugung der einen oder anderen Form zu erfahren. Solche Untersuchungen ließen sich auf diachrone Analysen zum „Lebenszyklus“ einer Form ausweiten. Das ist jedoch nicht Thema dieser Arbeit und soll daher auch hier nicht weiter behandelt werden.

⁶ zu Variations- und Invarianzbedürfnis siehe Köhler (1989:4, 8).

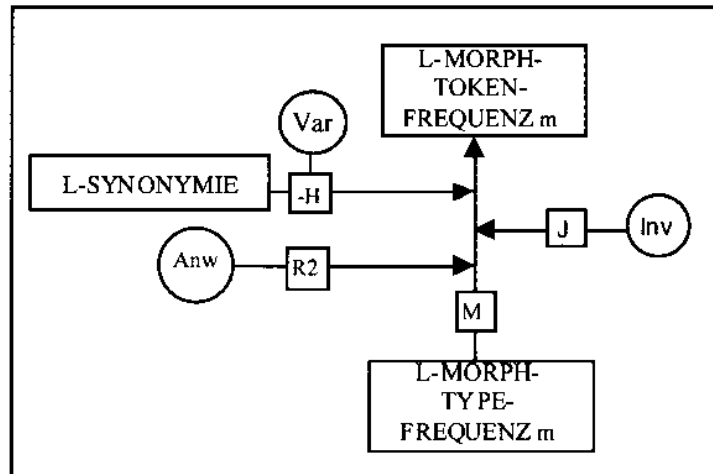


Abb. 2.1.5.3: Die Anhängigkeiten der *Morph-Token-Frequenz_m*

Die beiden Funktionen (vii) und (viii) lassen sich in einer Gleichung zusammenfassen. Dabei müssen jedoch sowohl das Anwendungs- als auch das Invarianzbedürfnis berücksichtigt werden. Nach Logarithmierung erhalten wir demnach:

$$L\text{-TOKEN-FREQU}_m = R_2 \cdot Anw - H \cdot L\text{-SYNONYMIE} + J \cdot Inv + M \cdot L\text{-TYPE-FREQU}_m$$

Mit graphischen Mitteln wird diese Formel in der Abbildung 2.1.5.3 noch einmal wiedergegeben.

2.1.6 Die Morphlänge

Um die Länge eines einzelnen Morphs zu bestimmen, wird analog zu den Zusammenhängen auf der lexikalischen Ebene (vgl. Köhler, 1986:69, 70) angenommen, daß eine Frequenz-Längen-Relation auch auf der morphologischen Ebene existiert, daß also jede Morphlänge einer Morph-Token-Frequenz entspricht.

Das Phänomen der Morph-Kürzung läßt sich besonders gut an Entlehnungen aufzeigen: Das Wort *Automobil* kann nicht in mehrere Morphe zerlegt werden, da es ohne Kenntnis der griechischen und lateinischen Sprache nicht möglich ist, die enthaltenen Konstituenten zu identifizieren. Daher können wir davon ausgehen, daß es sich um ein einmorphemiges Wort handelt. Mit zunehmender Relevanz des Automobils für die Moderne stieg die Verwendungshäufigkeit dieses Begriffs. Das Bedürfnis nach Minimierung des Produktionsaufwands (*minP*) ist dafür verantwortlich, daß das viersilbige Morph im Deutschen zu *Auto* gekürzt wurde; die Silbenlänge wurde demnach halbiert. In anderen Sprachen ist eine ähnliche Entwicklung zu beobachten: Das Indonesische kennt das Wort *mobil*; die Skandinavier benutzen sogar den einsilbigen Ausdruck *bil* (vgl. Köhler, 1986:69). Dem Bedürfnis nach Minimierung des Produktionsaufwands steht je-

doch das Bedürfnis nach Minimierung des Dekodierungsaufwands gegenüber. Denn Kürzungen von Morphen, die sich in der gesprochenen Sprache daran zeigen, daß aufgrund mangelndem Produktionsaufwand Endungen und ganze Morph- und Wortteile weggelassen bzw. ‚verschliffen‘ werden, erschweren das Verständnis.

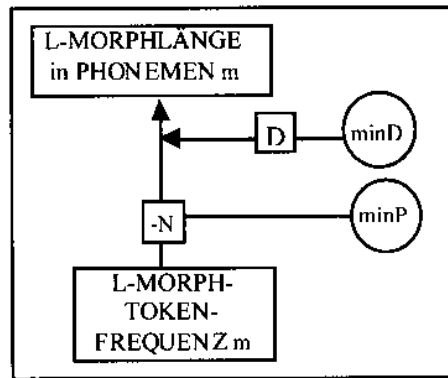


Abb. 2.1.6: Die Abhängigkeiten der individuellen Morphemlänge

Es kann auf der Grundlage obiger Ausführungen folgende Hypothese formuliert werden:

Hypothese 8: *Die Veränderungsrate der Morphemlänge ist umgekehrt proportional zur Morph-Token-Frequenz.*

Diese Aussage lautet in Form einer Differentialgleichung:

$$\frac{y'}{y} = \frac{-N}{x} \text{ mit der Lösung } y = C \cdot x^{-N}.$$

Ersetzt man die Variablen y durch die Morphemlänge und x durch die Morph-Token-Frequenz des Morphs m , so erhält man nach Logarithmierung die Formel

$$L\text{-MORPHLÄNGE}_m = \ln C - N \cdot L\text{-TOKEN-FREQU}_m$$

Die Konstante $\ln C$ läßt sich in diesem Fall bestimmen, indem man das Bedürfnis nach Minimierung des Dekodierungsaufwands berücksichtigt. Nach Ersetzung von $\ln C$ ergibt sich demnach die endgültige Funktion

$$L\text{-MORPHLÄNGE}_m = D \cdot \min D - N \cdot L\text{-TOKEN-FREQU}_m$$

Die Abbildung 2.1.6 soll diese Beziehung verdeutlichen.

2.2 Die Systemgrößen und ihre Funktionen

In der nachstehenden Tabelle sind die Systemgrößen, die im vorgestellten Modell enthalten sind, zusammen mit ihren Funktionen noch einmal aufgeführt.

Tabelle 1
Übersicht über die Systemgrößen und ihre Funktionen

Systemgröße	Funktion
Morph-Inventarumfang	$MIU = \text{Kod}^E \cdot \emptyset MP^{-S}$
Kombinatorisches Morph-Potential der Länge l	$KOM_l = MIU^l$
Restringiertes Morph-Kombinationspotential	$RMPOT = KOM \cdot SEM^{-U_1} \cdot ART^{-U_2} \cdot REZ^{-U_3} \cdot Red^{-U_4}$
Lexikalisiertes Morph-Kombinationspotential	$LEX_m = RMPOT_m \cdot \text{Kod}^K \cdot \text{minI}^{-l} \cdot MP_m^G$
Morph-LexType-Frequenz	$MLTF_m = LEX_m \cdot \text{Anw}^{R_1} \cdot SY^{-L}$
Morph-NeuType-Frequenz	$MNTF_m = POT_m \cdot \text{Kod}^V \cdot MSY_m^{-T}$
Token-Frequenz des Morphs m	$TOF_m = \text{Anw}^{R_2} \cdot SY^{-H} \cdot Inv^J \cdot TYF_m^M$
Morphlänge des Morphs m	$ML_m = \text{min D}^D \cdot TOF_m^{-N}$

Anw	:	Anwendungsbedürfnis
ART	:	Artikulatorische Restriktionen
Kod	:	Kodierungsbedürfnis
KOM	:	Kombinatorisches Morph-Potential (gesamt)
KOM _l	:	Kombinatorisches Morph-Potential der Länge l
LEX _m	:	Lexikalisiertes Morph-Kombinationspotential des Morphs m
MIU	:	Morph-Inventarumfang
minD	:	Bedürfnis nach Minimierung des Dekodierungsaufwands
minI	:	Bedürfnis nach Minimierung des Inventarumfangs
MLTF _m	:	Morph-LexType-Frequenz des Morphs m
MNTF _m	:	Morph-NeuType-Frequenz des Morphs m
ML _m	:	Länge des Morphs m
MSY _m	:	Synonymie des Morphs m
MP _m	:	Polysemie des Morphs m
∅MP	:	Polysemie des Morphs m
POT _m	:	Potentielle Neubildungen, die das Morph m enthalten
Red	:	Redundanzbedürfnis
REZ	:	Rezeptive Restriktionen

RMPOT	:	Restringiertes Morph-Kombinationspotential
RMPOT _m	:	Restringiertes Morph-Kombinationspotential des Morphs <i>m</i>
SEM	:	Semantische Restriktionen
SY	:	(Wort-)Synonymie
TOF _m	:	Token-Frequenz des Morphs <i>m</i>
TYF _m	:	Type-Frequenz des Morphs <i>m</i>

3. Die empirische Überprüfung des Modells

Das in den vorangehenden Kapiteln vorgestellte Modell setzt sich aus einer Reihe von (Gesetzes-)Hypothesen zusammen. Um dem Anspruch der Erklärungskraft zu genügen, ist es wichtig, daß die Hypothesen theoretische Plausibilität und Konsistenz aufweisen. Darüber hinaus ist es von großer Bedeutung, die Annahmen auch empirisch zu überprüfen. Wird eine Hypothese hinreichend an Datenmaterial getestet und aufgrund der Testergebnisse bestätigt, so kann sie letztendlich den Status eines Gesetzes erhalten.

Für einige der Hypothesen wurden entsprechende Stichproben erhoben. Dabei stand für fast alle der überprüften Beziehungen Datenmaterial aus drei Sprachen (Deutsch, Englisch und Niederländisch) zur Verfügung. Eine Beziehung, die nur am Deutschen getestet werden konnte, ist die Abhängigkeit des lexikalisierten Morph-Kombinationspotentials von der Morph-Polysemie.

Tabelle 2
Die empirisch überprüften Beziehungen des Modells

<i>Abhängigkeit</i>	<i>Funktion</i>
Lexikalisiertes Morph-Kombinationspotential des Morphs <i>m</i> als Funktion der Morph-Polysemie	$LEX_m = A \cdot MP_m^B$
Morph-LexType-Frequenz des Morphs <i>m</i> als Funktion des lexikalisierten Morph-Kombinationspotentials	$MLTF_m = A \cdot LEX_m$
Token-Frequenz des Morphs <i>m</i> als Funktion seiner Type-Frequenz	$TOF_m = A \cdot TYF_m^B$
Morphlänge des Morphs <i>m</i> als Funktion der Token-Frequenz	$ML_m = A \cdot TOF_m^B$

Tabelle 2 gibt die Beziehungen wieder, die empirisch überprüft wurden. Zur Morph-Type- und zur Morph-Token-Frequenz ist allerdings zu bemerken, daß bei den erhobenen Stichproben nur lexikalisierte Wörter berücksichtigt wurden. Im vorgestellten Modell wurde unter diesen Größen die Frequenz der lexikalisierten *und* der neugebildeten Wörter verstanden.

Im folgenden Teil dieser Arbeit wird auf das verwendete Datenmaterial näher eingegangen werden. Danach folgt eine detaillierte Darstellung der Testergebnisse obiger Untersuchungen.

3.1 Die Operationalisierung der Daten und das Datenmaterial

Eine empirische Überprüfung des aufgestellten Modells verlangt eine Festlegung, wie die theoretischen Größen identifiziert und gemessen werden sollen. Im folgenden werden deshalb für die Größen derjenigen Beziehungen, die empirisch untersucht werden sollen, Meßvorschriften angegeben.

Für die Überprüfung wurde als Datenmaterial im wesentlichen die CD-ROM-Version der lexikalischen Datenbank von CELEX⁷ verwendet, die aus einem deutschen, englischen und niederländischen Teil besteht. Das deutsche Lexikon enthält 50708 Lemmata, das englische 52447 und das niederländische 124136. Auf der Basis der morphologischen Zerlegungen dieser Lemmata konnten für die deutsche Sprache eine Liste von 18097 Morphen erstellt werden. Für das Englische ergaben sich 17513, für das Niederländische 31627 Morphe⁸.

Auf der Grundlage der CELEX-Lexika wurden für die folgenden Größen Werte bestimmt:

- Morphinventar
- Wortinventar
- Lexikalisiertes Morph-Kombinationspotential
- Morph-LexType-Frequenz
- Morph-Token-Frequenz
- Morphiänge

In einzelnen wurde bei der Bestimmung der Werte folgendermaßen vorgegangen:

Morphinventar und Wortinventar

Das Morphinventar wurde mittels der morphologischen Zerlegung der Lemmata aufgebaut. Es entspricht den Morphen, die in den Lemmata von CELEX enthalten sind. Als Wortinventar wurden bei allen Untersuchungen die Lemmata von CELEX verwendet. Es wurde also in keinem Fall auf Wortformen oder andere Wörter zurückgegriffen.

Lexikalisiertes Morph-Kombinationspotential

Der Wert des *lexikalisierten Kombinationspotentials* eines Morphs wurde bestimmt über die Anzahl der Lemmata in CELEX, die das Morph enthalten. Dafür

⁷ Genaue Quellenangabe siehe Baayen etc. (1993).

⁸ In CELEX sind einigen Lemmata mehrere Zerlegungen zugeordnet. Für die Erstellung der Morphlisten wurde grundsätzlich die erste Zerlegung berücksichtigt.

wurden die in CELEX enthaltenen morphologischen Zerlegungen der Lemmata verwendet.

Morph-LexType-Frequenz und Morph-Token-Frequenz

Für die *Morph-LexType-Frequenz* und die *Token-Frequenz* wurden die Frequenzangaben verwendet, die in CELEX den Lemmata zugeordnet sind. Dabei entspricht die *Morph-LexType-Frequenz* (bzw. *Morph-Type-Frequenz*) eines Morphs *m* der Anzahl aller Lemmata, die das Morph enthalten und eine Frequenz größer als Null aufweisen. Die Summe der Einzelfrequenzen dieser Lemmata wurde als Wert für die *Token-Frequenz* des Morphs angesetzt. Streng genommen handelt es sich dabei nur um einen Näherungswert. Denn die Größen ‚*Morph-Token-Frequenz*‘ und ‚*Morph-Type-Frequenz*‘ sind im dargestellten Modell so zu verstehen, daß mit ihnen nicht nur die Häufigkeit der lexikalisierten Wörter, sondern auch der neugebildeten Wörter, die das Morph *m* enthalten, gemeint ist.

Die Morphemlänge

Die Morphemlänge läßt sich theoretisch auf drei verschiedene Arten messen: In Graphemen, Phonemen und Silben. Da in CELEX nur die phonologische Information der Lemmata, nicht aber der Morphe enthalten ist, wurde die *Morphlänge* nur als Anzahl der Grapheme ermittelt, aus denen sich das jeweilige Morph zusammensetzt.

Die Morph-Polysemie

Die bisher aufgeführten Systemgrößen konnten alle mit Hilfe von CELEX bestimmt werden. Für die Festlegung der Bedeutungsanzahl eines Morphs mußte jedoch eine andere Lösung gefunden werden. Allgemein stellt diese Messung ein besonders schwieriges Problem dar: Für die Bestimmung der Polysemie von Wörtern lassen sich Bedeutungswörterbücher heranziehen, in denen zu jedem aufgeführten Lemma eine oder mehrere Bedeutungen aufgeführt sind; für Morphe bzw. Morpheme ist das jedoch problematischer. Nur im Falle der Wortbildungsmorpheme läßt sich aus der Wortbildungsliteratur zu jedem Morphem eine oder mehrere Funktionen (Bedeutungen) entnehmen. Die Anzahl dieser Funktionen kann als Maß für die Morph-Polysemie gewertet werden.

Um die Polysemie der Grundmorpheme zu bestimmen, wurde folgende Methode angewendet: Jedes Morph läßt sich mit einem oder mehreren Simplizia eines Bedeutungswörterbuchs identifizieren. Bei diesen Simplizia ist zu entscheiden, ob und welche Wörter von einem anderen abgeleitet wurden. Dadurch erhält man zu jedem Morph eine Restmenge an Simplizia, die von ihrer Wortbildungsgeschichte her nicht mehr zusammenhängen. Um einen Wert für die Morph-Polysemie eines bestimmten Morphs zu erhalten, muß dann die Bedeutungsanzahl jedes dieser Wörter dem Wörterbuch entnommen und daraus eine Summe gebildet werden.

Für die Bestimmung der Morph-Polysemie der deutschen Grundmorpheme (die englischen und niederländischen Morphe wurden hinsichtlich ihrer Bedeutungen nicht untersucht) wurde Datenmaterial verwendet, das auf der Grundlage des Wahrig - Deutsches Wörterbuch innerhalb des Projekts ‚Sprachliche Synergetik‘ (Prof. Dr. G. Altmann, Ruhr-Universität Bochum) erstellt und im Fach Linguistische Datenverarbeitung der Universität Trier weiter aufbereitet worden ist. Diese Datenbasis enthält 96.601 Lemmata, von denen 10.487 durch eine ‚Bedeutungshierarchie‘ ergänzt wurden, welche den entsprechenden Einträgen des an der Akademie der Wissenschaften der DDR erarbeiteten und von Kempcke (1984) herausgegebenen Handwörterbuch der deutschen Gegenwartssprache entnommen wurden. Es handelt sich bei der Bedeutungshierarchie eines Lemmas um die Auflistung der römischen und arabischen Ziffern, die den entsprechenden Eintrag im Handwörterbuch der deutschen Gegenwartssprache in seine verschiedenen Bedeutungen und Unterbedeutungen zergliedert. Die Anzahl dieser Ziffern wurde als Bedeutungsanzahl des betreffenden Lemmas gewertet.

Zur Bestimmung der Polysemie eines Grundmorphems wurde nun nach obigem Verfahren auf die Bedeutungsangaben der Simplizia im ‚Wahrig-Lexikon‘ zurückgegriffen, die sich mit dem betreffenden Morph identifizieren lassen. Dabei wurde berücksichtigt, daß den Morphen in den morphologischen Zerlegungen der Lemmata in CELEX immer ihre ursprüngliche Wortart zugewiesen wird. Daher wurden nur die Simplizia berücksichtigt, welche die entsprechende Wortart des Morphs aufwiesen. So läßt sich beispielsweise das Morph *aal* mit den beiden Simplizia *Aal* und *aalen* identifizieren. Beide Wörter kämen also zur Bestimmung der Morph-Polysemie von *aal* in Frage. Die morphologische Zerlegung in CELEX zeigt jedoch, daß das Verb *aalen* vom Substantiv *Aal* abgeleitet wurde. Daher wurde als Polysemie-Wert von *aal* nur die Bedeutungsanzahl des Substantivs *Aal* herangezogen. Im Falle des Morphs *lauf* dagegen läßt sich CELEX entnehmen, daß das Substantiv *Lauf* vom Verb *laufen* abgeleitet wurde. Daher wurde in diesem Fall die Bedeutungsanzahl von *lauf* mit der Bedeutungsanzahl des Verbs *laufen* gleichgesetzt.

Nach Berücksichtigung der Ableitungsgeschichte kam es vor, daß zwei oder mehrere Simplizia zu einem Morph gefunden werden, die nicht voneinander abgeleitet werden können. In diesen Fällen wurden die Bedeutungszahlen der Wörter einfach summiert.

Die Polysemie-Werte der Wortbildungsmorpheme, die im Deutschen zu den Affixen gehören, wurden auf zwei verschiedenen Wegen bestimmt:

1. Im Morphemregister zu Band I-III des mehrbändigen Werks Deutsche Wortbildung: Typen und Tendenzen in der Gegenwartssprache von Kühnhold und Prell (im folgenden mit ‚KuP-Affixliste‘ bezeichnet) werden die verschiedenen Bedeutungen eines Morphems mittels Indizes notiert. Die Anzahl dieser Indizes diente als Wert für die Polysemie des Affixes.

2. Da gerade die Bedeutungen der Affixe sehr schwer zu bestimmen sind, wurde zur Kontrolle eine von H. Bluhme aufgestellte Affixliste („BLUHME-Affixliste“) für eine zweite Messung herangezogen. Der Polysemie-Wert konnte über die Anzahl der dort aufgeführten möglichen Bedeutungen eines Affixes bestimmt werden.

3.2 Die Testmethoden

Unter Zuhilfenahme der empirisch gewonnenen Daten können die konstatierten Abhängigkeiten, die sich aus dem vorgestellten Modell ableiten lassen, überprüft werden. Dafür werden entsprechende Daten erhoben und Regressionsanalysen durchgeführt, welche die Funktionsgleichungen an die empirischen Daten anpassen. Die Güte der Anpassung kann in einem weiteren Schritt mit Hilfe des Determinationskoeffizienten R^2 gemessen werden.

Da die aufgestellten Beziehungen im Grunde nichtlineare Funktionen darstellen, bietet sich eine nichtlineare Regression an. Auf der anderen Seite ist es jedoch auch möglich, auf der Grundlage der linearisierten Funktionsgleichungen, die auch schon für die graphische Darstellung des Modells verwendet wurden, eine lineare Regression durchzuführen. Lineare und nichtlineare Regression bringen jeweils unterschiedliche Vor- und Nachteile mit sich (vgl. Giesecking, 1993). Es wurde entschieden, die Funktionen des Modells zu linearisieren und mittels einer linearen Regression anzupassen.

Zusätzlich zur Regression wird im Falle des linearisierten Modells mittels des t -Tests getestet, ob zwischen den jeweiligen Variablen ein signifikanter Zusammenhang besteht. Es kann allerdings schon vorweggenommen werden, daß die Ergebnisse des t -Tests für die untersuchten Daten keine so große Aussagekraft besitzen, wie es wünschenswert wäre. Denn dieser Test verliert bei großen Datenmengen, wie sie auch für die Untersuchungen dieser Arbeit vorliegen, an Signifikanz. Da jedoch der t -Test nicht immer einen linearen Zusammenhang der jeweils untersuchten Systemgrößen bestätigt hat, werden seine Ergebnisse dennoch aufgeführt.

Um Verfälschungen des Determinationskoeffizienten entgegenzuwirken, der durch die Mittelwertbildung ganz unterschiedlicher Anzahl von y -Werten entstehen kann, werden, wie Grotjahn (1992:150) es vorschlägt, die (x, \bar{y}) -Paare für die vorgenommenen Untersuchungen gewichtet. Allerdings wird dafür nicht der von ihm vorgeschlagene modifizierte Determinationskoeffizient verwendet (vgl. Grotjahn, 1992:162), sondern jedes Wertepaar erhält als Gewicht die Anzahl der y -Werte, aus denen der Mittelwert \bar{y} gebildet wurde. Die Regressionsanalyse wird dann unter Berücksichtigung dieses Gewichts vorgenommen. Als Ergebnis erhält man eine Kurve, die sich denjenigen Datenpunkten, die den Großteil aller Daten ausmachen, besonders gut annähert. In dünn besetzten Gebieten kann sie jedoch erheblich von den Datenpunkten abweichen.

3.3 Einzelergebnisse der Untersuchung

3.3.1 Die Abhängigkeit des lexikalisierten Morph-Kombinationspotentials von der Morph-Polysemie

Aus der Struktur des Modells läßt sich die Hypothese ableiten, daß das lexikalisierte Morph-Kombinationspotential eines Morphs m eine Funktion seiner Morph-Polysemie ist. Diese Hypothese wurde in der nichtlinearisierten Form durch die Gleichung $LEX_m = A \cdot MP_m^B$ ausgedrückt (siehe Hypothese 3). Dabei steht die Konstante A für den kombinierten Einfluß des restringierten Morph-Kombinationspotentials und des Gleichgewichts zwischen den Bedürfnissen Kod und $minI$; die Konstante B gibt die Einflußstärke der Morph-Polysemie wieder⁹. Die Anpassung der linearisierten Funktion an die Daten der Stichproben wurde, wie schon vorweggenommen wurde, mit gewichteten Datenpaaren durchgeführt. Wie wichtig diese Vorgehensweise in diesem Fall ist, wird an der Häufigkeitsverteilung der Morph-Polysemiewerte deutlich.

Die Untersuchung der Stichprobe, die auf der Grundlage des Wahrig-Lexikons und der KuP-Affixliste erhoben wurde, ergab eine Streuung der Morph-Polysemie zwischen den Werten 1 und 37. Wie die Abb. 1 zeigt, verteilen sich 95% der Morphe auf die Morph-Polysemie-Werte 1 bis 8, das entspricht ca. 24% der gefundenen Werte¹⁰. Für die BLUHME-Affixliste gelten dieselben Ergebnisse. Daher würde eine Regressionsanalyse ohne Gewichtung der Daten zu einer erheblichen Beeinträchtigung des Determinationskoeffizienten führen. Die Anpassung der linearisierten Funktion an die gewichteten Daten der Stichproben führte zu den Ergebnissen der Tabelle 3.

Tabelle 3
Anpassung der Funktion $L-LEX_m = \ln A + B \cdot L-MP_m$

<i>Sprache</i>	<i>verwendete Affixliste</i>	<i>A</i>	<i>B</i>	<i>R²</i>	<i>Sign. t</i>
Deutsch	KuP	3,22460	1,15072	0,65107	0,0000
Deutsch	BLUHME	4,35956	0,98861	0,54434	0,0000

Die Determinationskoeffizienten (R^2) weisen darauf hin, daß es sich nur um mäßig gute Anpassungen handelt. Aber das Signifikanzniveau von t , das in beiden

⁹ Die folgenden Untersuchungsergebnisse werden teilweise auch in Krott (1998) diskutiert. Allerdings wird dort im Gegensatz zu den hier dargestellten Untersuchungen nur über die KuP-Affixliste Polysemiewerte für Affixe erhoben. Auf der anderen Seite wird in Krott (1989) detaillierter auf die Probleme der Funktionswörter eingegangen.

¹⁰ Die Abbildung zeigt nur einen Ausschnitt der Verteilung. Die fehlenden Werte (ab einem Polysemie-Wert von 16) liegen jedoch alle unter 0,5%.

Fällen unter 0,00005 liegt, zeigt an, daß die Größen der logarithmierten Gleichung dennoch in einer linearen Beziehung zueinander zu stehen scheinen.

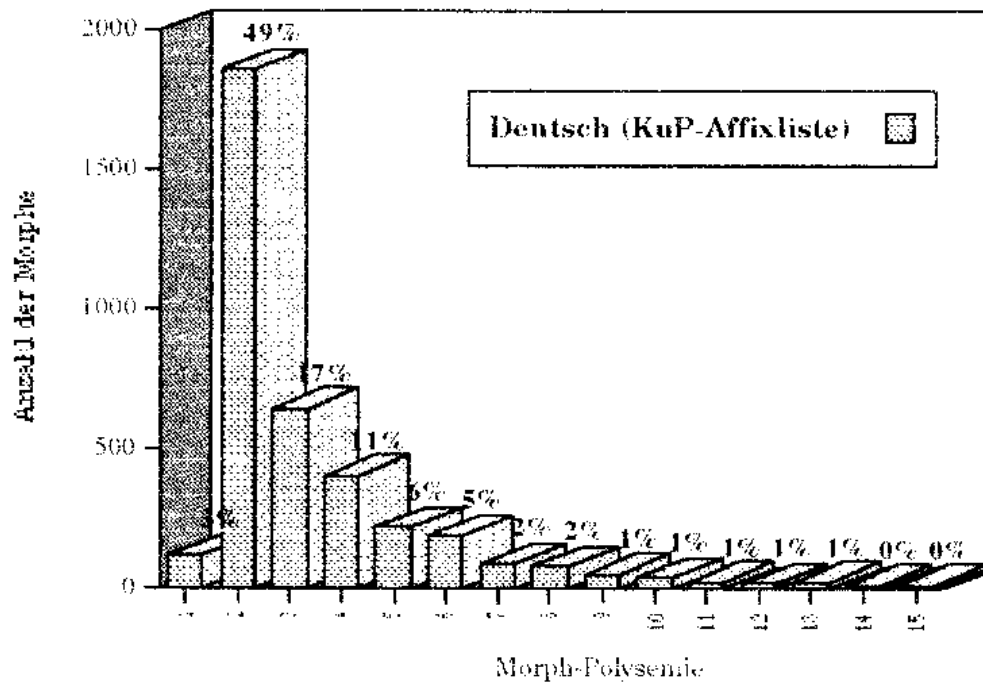


Abb. 1: Häufigkeitsverteilung der Morph-Polysemie (Ausschnitt)

Eine bestehende Abhängigkeit läßt auch der optische Eindruck der Abb. 2 und 3 vermuten, welche die Daten der Stichproben und die theoretischen, *nicht-linearen* Funktionen mit den aus der Stichprobe geschätzten Parametern zeigen. Besonders im Bereich der kleinen Morph-Polysemie-Werte ($MP < 9$), der, wie wir gesehen haben, den größten Teil der Daten ausmacht, folgen die Funktionen eng den Datenpunkten. Allerdings ist auffallend, daß gerade die Morphe mit einer einzigen Bedeutung überdurchschnittlich viele Vorkommen im Lexikon aufweisen. Dieser hohe Wert ist auf eine Reihe von Affixen zurückzuführen, die genau eine Bedeutung besitzen und eine sehr hohe Frequenz im Lexikon aufweisen (z. B. *-s*: 1814 Vorkommen; *-ier*: 915 Vorkommen; *-keit*: 801 Vorkommen). Diese Produktivität kann auf die Transparenz dieser Affixe zurückgeführt werden.

Da der vermutete Zusammenhang zwischen den Größen anscheinend doch besteht, stellt sich die Frage, warum die Determinationskoeffizienten nur bei 0,54 bzw. 0,65 liegen. Eine mögliche Erklärung dafür wäre, daß die Daten keine Homogenität aufweisen, da es sich um verschiedene Typen von Morphen handelt¹¹. Ein Indiz dafür ist der zu große Wert für das Vorkommen der Morphe mit einer

¹¹ Zum Problem der Datenhomogenität siehe Altmann (1992).

einigen Bedeutung, der auf einige Affixe zurückgeführt werden kann. Eine andere Möglichkeit wäre, daß die Messung der Systemgröße Morph-Polysemie einige Mängel aufweist. Die Bestimmung der Bedeutungsanzahl über die Polysemie der Simplizia birgt eine mögliche Fehlerquelle in sich. Denn bei der Auswahl eines bestimmten Simpliziums ist es u.U. möglich, daß Bedeutungen vernachlässigt werden, die andere, nicht jedoch das gewählte Wort besitzen. Dieses Problem kann allerdings mit den zur Verfügung stehenden Mitteln nur unter großem Aufwand gelöst werden, der hier nicht geleistet werden kann.

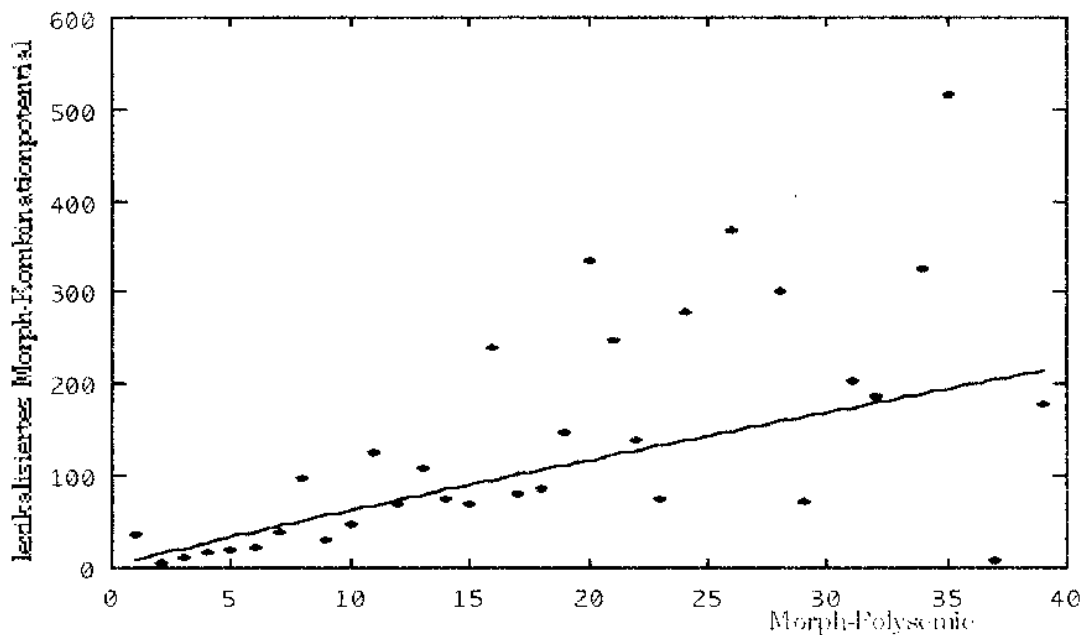


Abb. 2: Die Abhängigkeit des lexikalisierten Kombinations-Potentials von der Morph-Polysemie mit Berücksichtigung der KuP-Affixliste (Deutsch)

Außer diesem Problem muß festgehalten werden, daß die Bestimmung der Polysemie von Affixen und derjenigen Morphe, die als Funktionswörter verwendet werden, äußerst problematisch ist, besonders deshalb, weil sie gegenüber den Bedeutungswortklassen - wie der Name schon sagt - in erster Linie Funktionen erfüllen, die sich erheblich von denen der Nomen, Verben oder Adjektive unterscheiden. Aus diesem Grund ist es sinnvoll, die unterschiedlichen Morphtypen getrennt voneinander zu untersuchen. Dafür werden als Typen die ursprünglichen Wortarten der Morphe, die sich den morphologischen Zerlegungen in CELEX entnehmen lassen, und der Morphtyp ‚Affix‘ betrachtet.

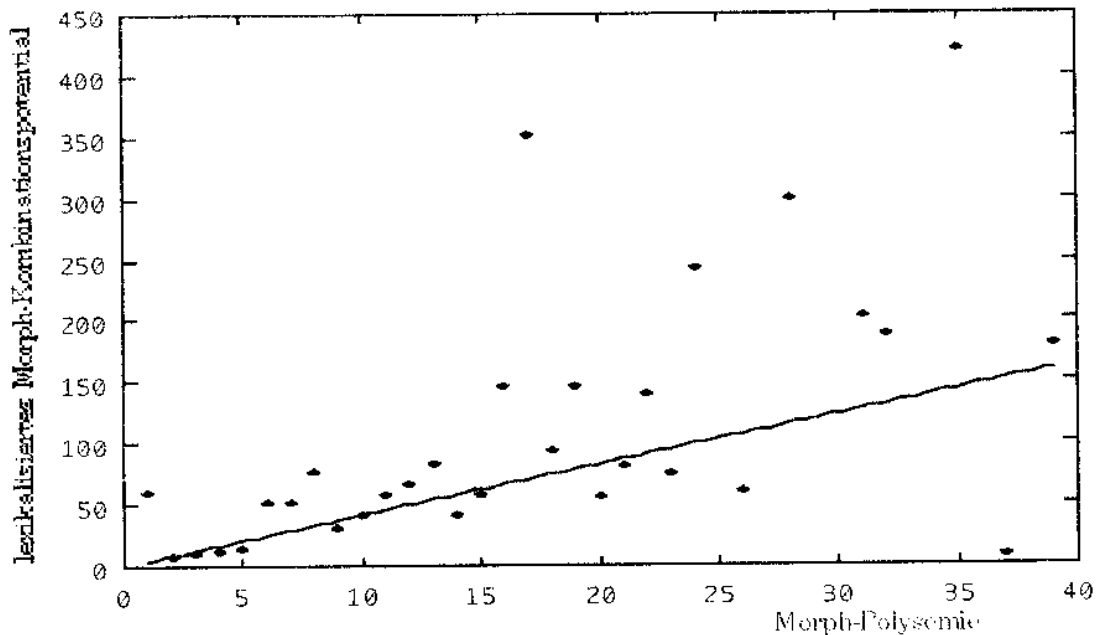


Abb. 3: Die Abhängigkeit des lexikalisierten Kombinations-Potentials von der Morph-Polysemie mit Berücksichtigung der BLUHME-Affixliste (Deutsch)

Tabelle 4 zeigt die Ergebnisse dieser Untersuchungen. Da die Morphtypen ‚Numerale‘, ‚Artikel‘, ‚Interjektion‘, ‚Stamm‘ und ‚flektierte Form‘ in der Stichprobe zu selten vertreten sind, konnten für diese Morphtypen keine Anpassungen durchgeführt werden.

Die Anpassungen der drei Morphtypen ‚Nomen‘, ‚Adjektiv‘ und ‚Verb‘ weisen gute bis sehr gute Determinationskoeffizienten auf, und das Signifikanzniveau von t bestätigt in diesen Fällen die lineare Abhängigkeit der Größen in der logarithmierten Form der Funktionsgleichung.

Die Morphtypen ‚Pronomen‘, ‚Präposition‘, ‚Konjunktion‘ und ‚Adverb‘ weisen dagegen schlechte Anpassungen auf. Das wird auch durch den t -Test unterstützt, so daß nicht von einer (linearen) Beziehung ausgegangen werden sollte.

Diese Ergebnisse unterstützen die obige Vermutung, daß die Morphe, die als Funktionswörter verwendet werden, das Gesamtergebnis verschlechtern, da die Bestimmung ihrer Bedeutung große Probleme aufweist. Es ist sogar fraglich, ob überhaupt von Bedeutungen dieser Morphe gesprochen werden sollte, zumindest in dem Sinne, wie Bedeutung bei Nomen, Adjektiven oder Verben verstanden wird. Diese Fragen zu klären, würde hier zu weit führen. Es muß aber festgehalten werden, daß über die Bestimmung der Bedeutungs- bzw. Funktionsanzahl der Funktionswörter noch einmal nachgedacht werden muß. Das kann auch dazu führen, daß diese Wörter bzw. Morphe aus der betrachteten Beziehung herausgenommen werden dürfen bzw. müssen.

Tabelle 4

Anpassung der Funktion $L-LEX_m = \ln A + B \cdot L-MP_m$ für die einzelnen Morphtypen

<i>Morphtyp</i>	<i>Sprache</i>	<i>A</i>	<i>B</i>	<i>R²</i>	<i>Sign. t</i>
Nomen	Deutsch	4,0029	0,98843	0,92105	0,0000
Adjektiv	Deutsch	3,75433	0,92173	0,79888	0,0000
Verb	Deutsch	4,03457	1,01888	0,95774	0,0000
Pronomen	Deutsch	7,60899	-0,1066	0,00959	0,7621
Adverb	Deutsch	9,15024	0,21087	0,02513	0,5726
Präposition	Deutsch	10,8281	0,82224	0,28052	0,0238
Konjunktion	Deutsch	1,79472	-0,0630	0,00935	0,7904
Affix (KuP-Liste)	Deutsch	45,2013	0,80148	0,67289	0,0000
Affix (BLUHME-Liste)	Deutsch	98,0169	1,06791	0,82570	0,0018

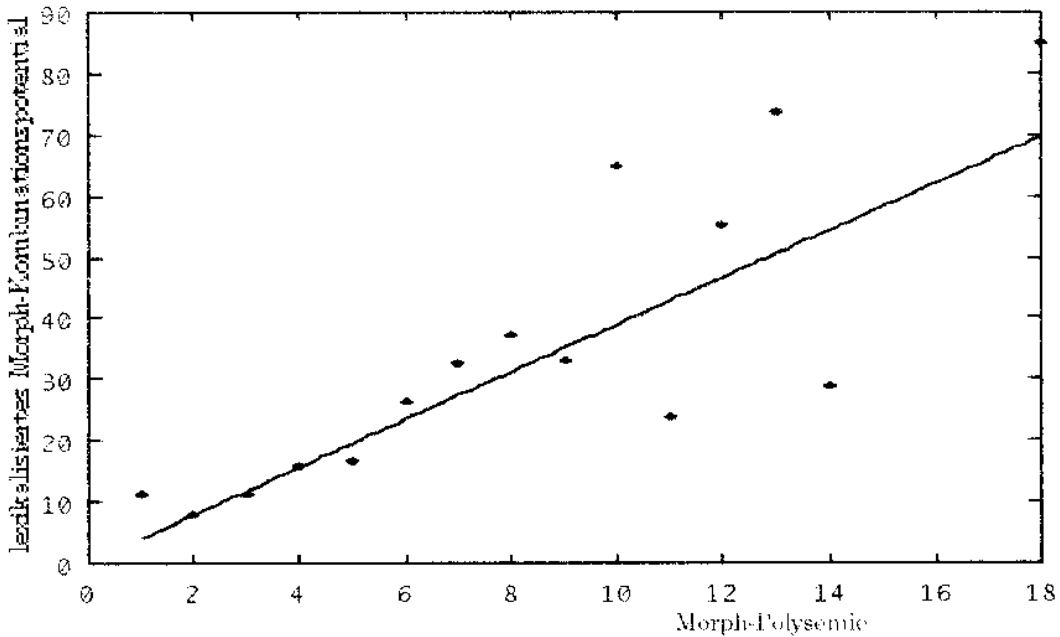


Abb. 4: Die Abhängigkeit des lexikalisierten Kombinations-Potentials von der Morph-Polysemie: Morphtyp ‚Nomen‘ (Deutsch)

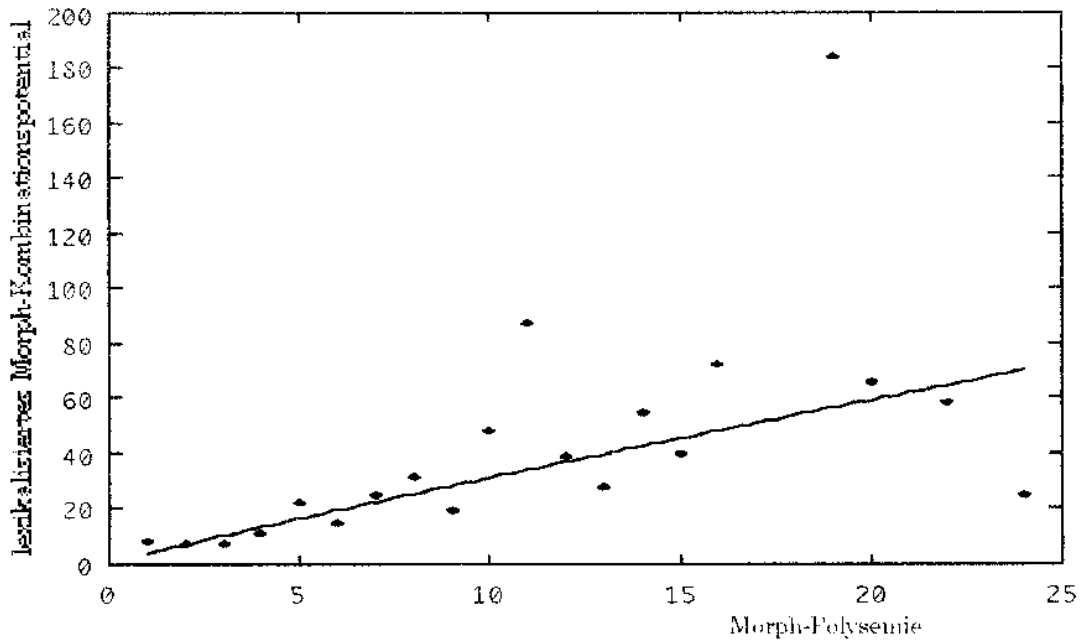


Abb. 5: Die Abhängigkeit des lexikalisierten Kombinations-Potentials von der Morph-Polysemie: Morphtyp ‚Adjektiv‘ (Deutsch)

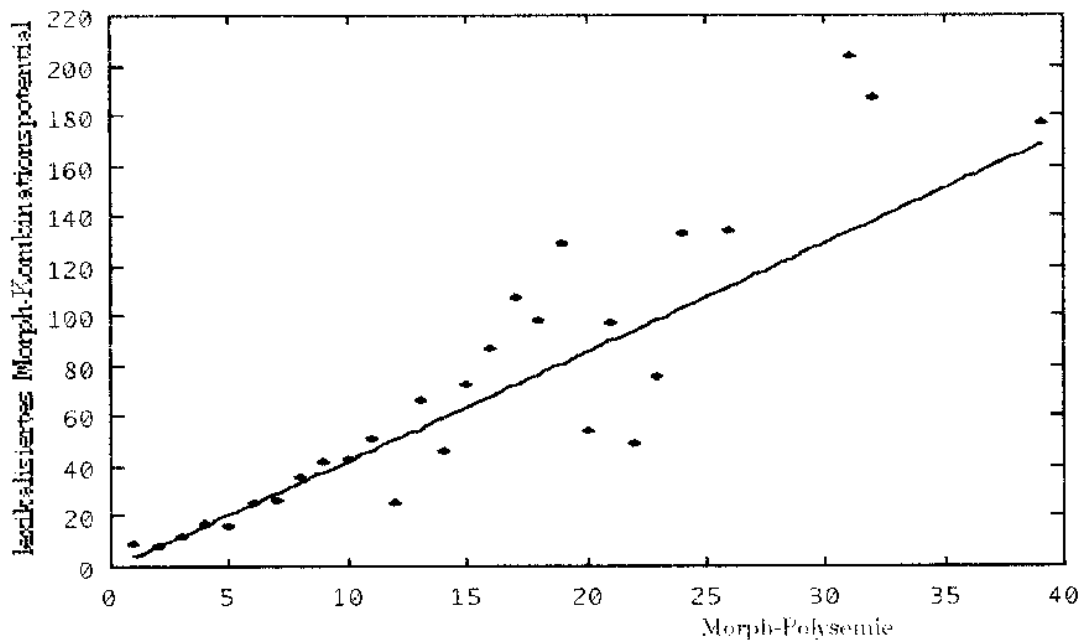


Abb. 6: Die Abhängigkeit des lexikalisierten Kombinations-Potentials von der Morph-Polysemie: Morphtyp ‚Verb‘ (Deutsch)

Die Ergebnisse der Anpassungen an die Daten der beiden Affixlisten unterstützen die aufgestellte Hypothese über die Abhängigkeit des lexikalisierten Morph-Kombinationspotentials von der Polysemie. Allerdings ist der Wert von R^2 bei der Anpassung an die Daten der BLUHME-Liste weitaus besser als im Falle der KuP-Affixliste. Aber auch in diesem Fall kann die Güte der Anpassung noch als Bestätigung der Hypothese angesehen werden, die allerdings verhältnismäßig schwach ist. Die Ergebnisse des t -Tests zeigen in beiden Fällen, daß es sich nach Linearisierung des Modells tatsächlich um eine lineare Abhängigkeit der Größen handelt. Der t -Test besitzt in diesen beiden Fällen eine höhere Signifikanz als bei der Betrachtung der Gesamtmenge der Morphe, da die Datenmengen mit 183 (KuP-Affixliste) bzw. 216 (BLUHME-Affixliste) verschiedenen Affixen relativ klein sind.

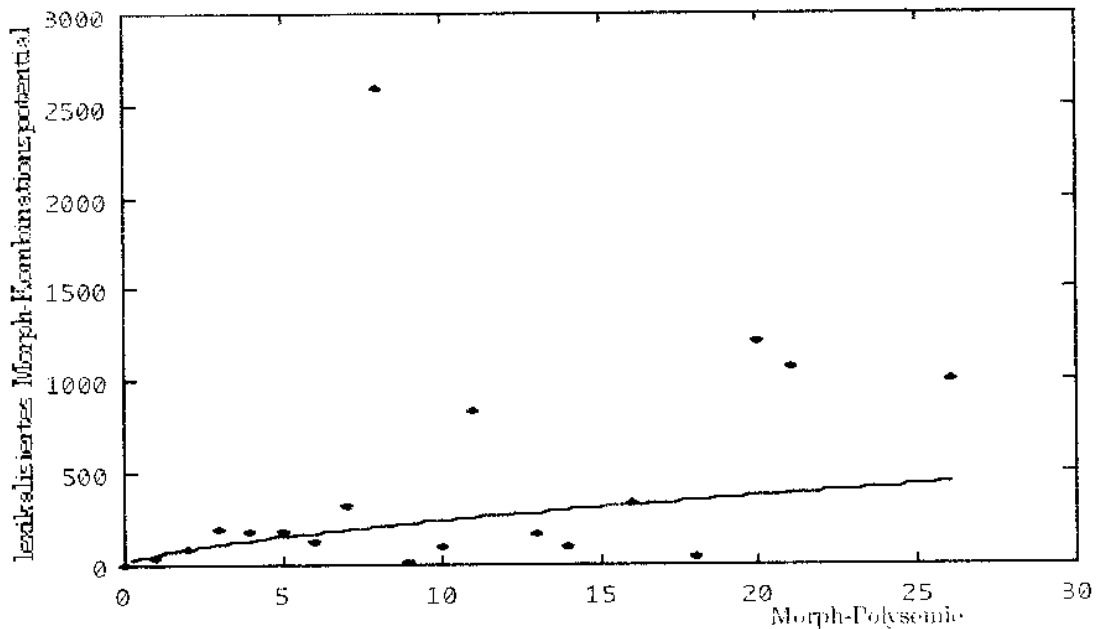


Abb. 7: Die Abhängigkeit des lexikalisierten Kombinations-Potentials von der Morph-Polysemie: Affixe der KuP-Affixliste (Deutsch)

Die Abb. 7 und 8 lassen die Unterschiede der beiden Stichproben erkennen. Abb. 7, welche die Daten der KuP-Affixliste zeigt, weist eine weitaus größere Bandbreite von Polysemie-Werten auf als Abb. 8. Daß trotz der relativ hohen Werte von R^2 im Bereich größerer Polysemie-Werte erhebliche Abweichungen der Datenpunkte bestehen, ist auf die Gewichtung der Daten bei der Regression zurückzuführen.

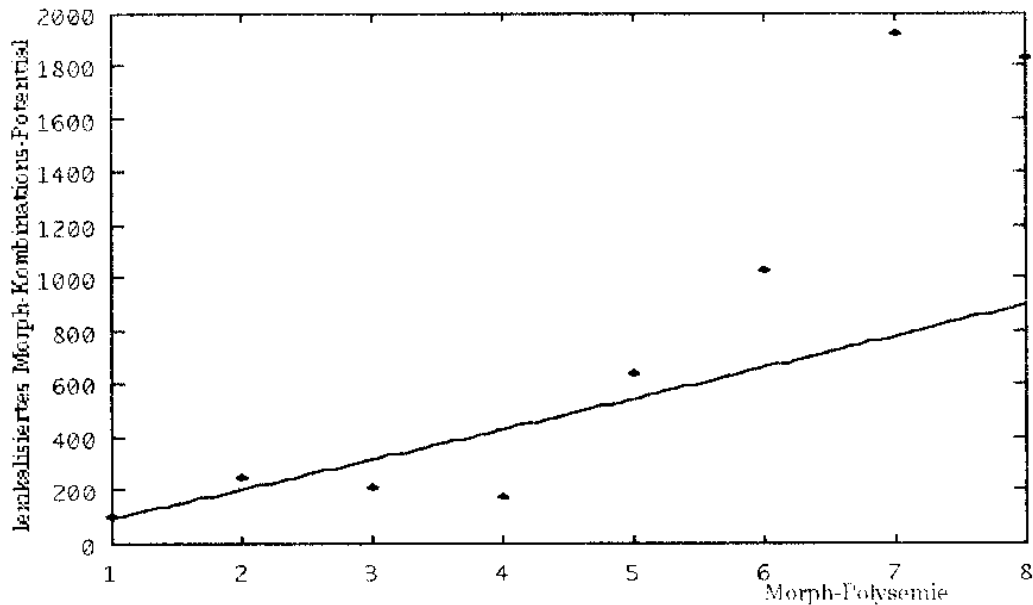


Abb. 8: Die Abhängigkeit des lexikalisierten Kombinations-Potentials von der Morph-Polysemie: Affixe der BLUHME-Affixliste (Deutsch)

3.3.2 Die Abhängigkeit der Morph-LexType-Frequenz vom lexikalisierten Morph-Kombinationspotential

Eine weitere Hypothese, die sich aus der Struktur des Modells herleitet, besagt, daß die Morph-LexType-Frequenz eines Morphs m eine Teilmenge seines lexikalisierten Morph-Kombinationspotentials darstellt, und daß diese Teilmenge mit der Größe des lexikalisierten Morph-Kombinationspotentials steigt.

Die zugehörige Funktionsgleichung lautet $MLTF_m = A \cdot LEX_m$, wobei die Konstante A eine Kombination der Einflüsse darstellt, die das Anwendungsbedürfnis und die Anzahl synonyme Wörter ausüben.

Wie aus Abb. 9 zu ersehen ist, sind die Daten aller drei Sprachen so ungleichmäßig verteilt, daß die Anpassung der theoretischen Funktion wieder mit gewichteten Datenpaaren vorgenommen wurde. Die Ergebnisse dieser Untersuchungen werden in der folgenden Tabelle wiedergegeben.

Wie an den Werten von R^2 zu erkennen ist, sind alle drei Anpassungen sehr gut. Das wird auch durch die Abb. 10 bis 12 deutlich, welche die Daten und die theoretischen Funktionen mit den aus den Daten geschätzten Parametern wiedergeben. Die aufgestellte Hypothese konnte demnach bestätigt werden.

Tabelle 5

Anpassung der Funktion $MLTF_m = A \cdot LEX_m$

<i>Sprache</i>	<i>A</i>	<i>R²</i>
Deutsch	0,70847	0,99257
Englisch	0,72840	0,97742
Niederländisch	0,62706	0,98718

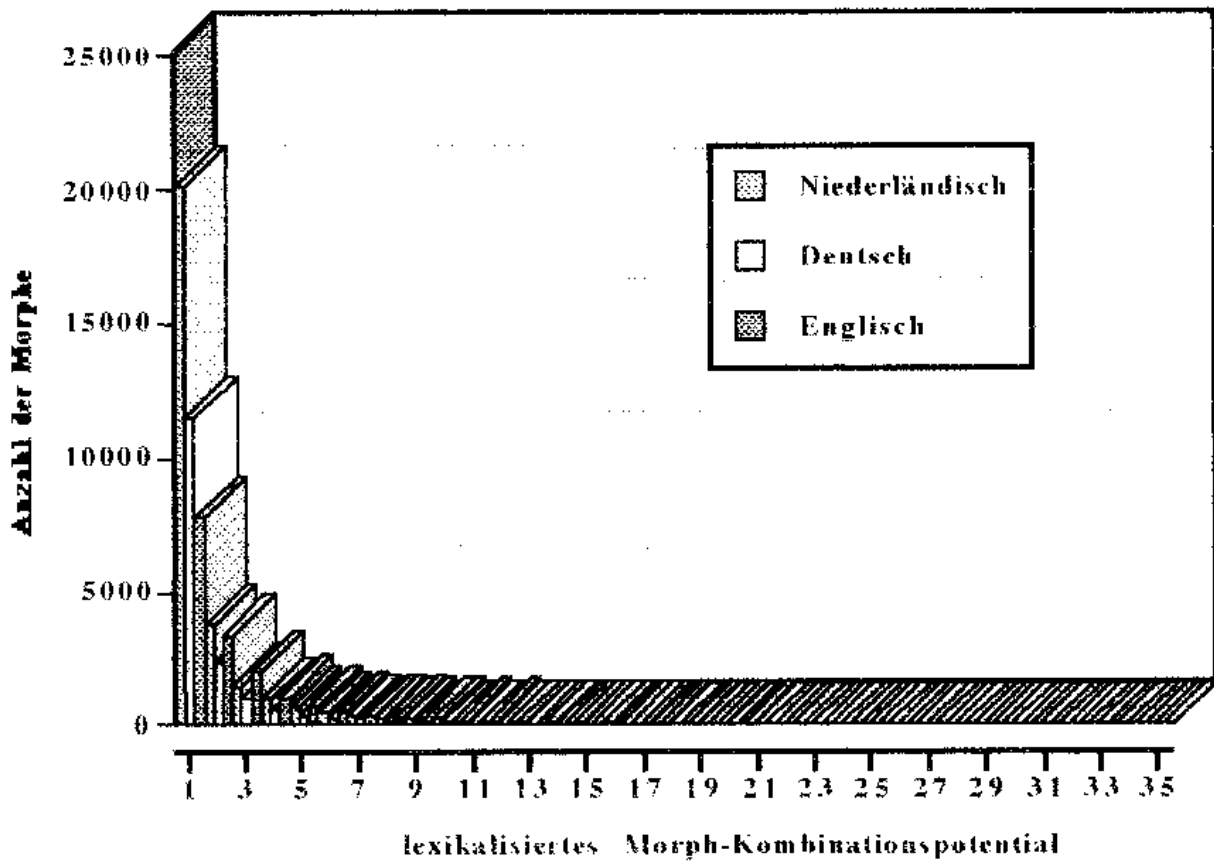


Abb. 9: Häufigkeitsverteilung des lexikalisierten Morph-Kombinationspotentials (Ausschnitte)

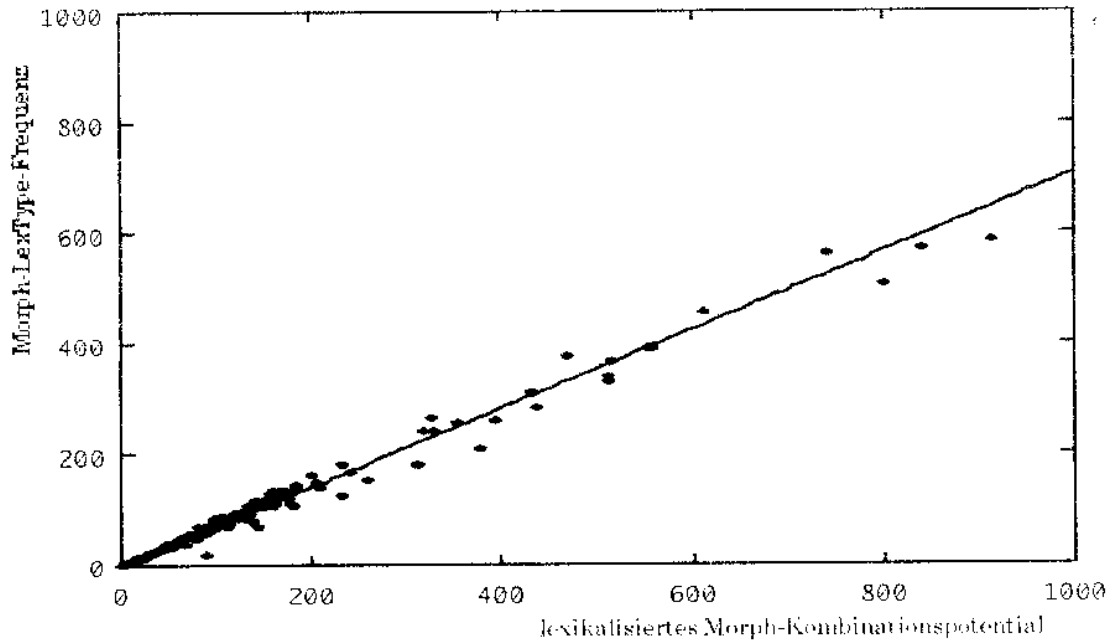


Abb. 10: Die Abhängigkeit der Morph-LexType-Frequenz vom lexikalisierten Morph-Kombinationspotential (Deutsch) (Ausschnitt)

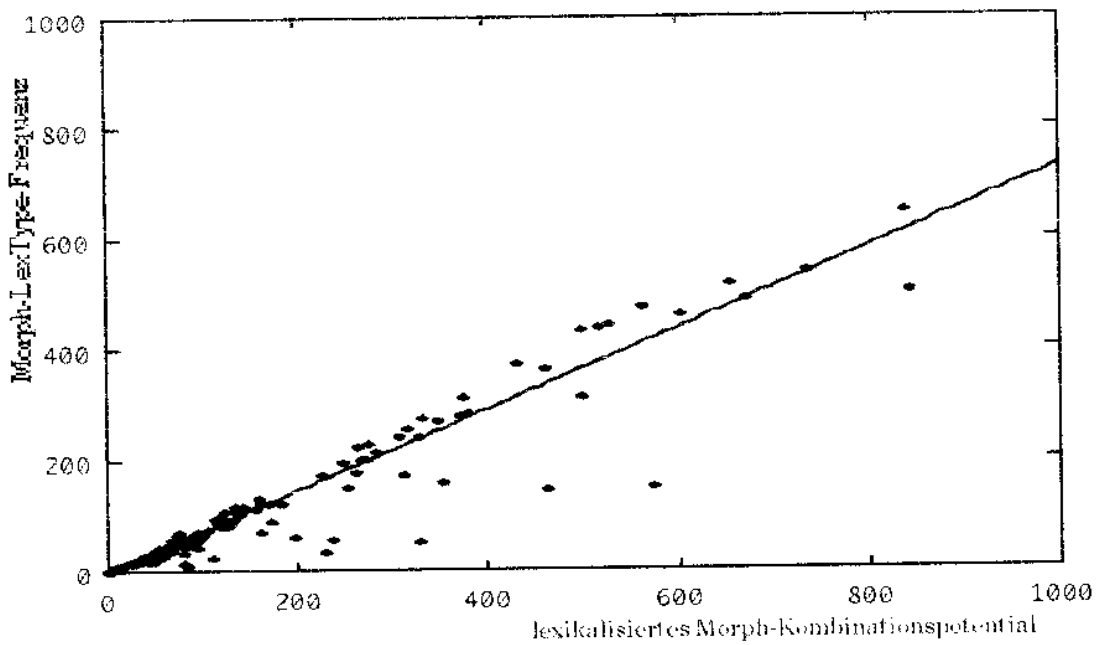


Abb. 11: Die Abhängigkeit der Morph-LexType-Frequenz vom lexikalisierten Morph-Kombinationspotential (Englisch) (Ausschnitt)

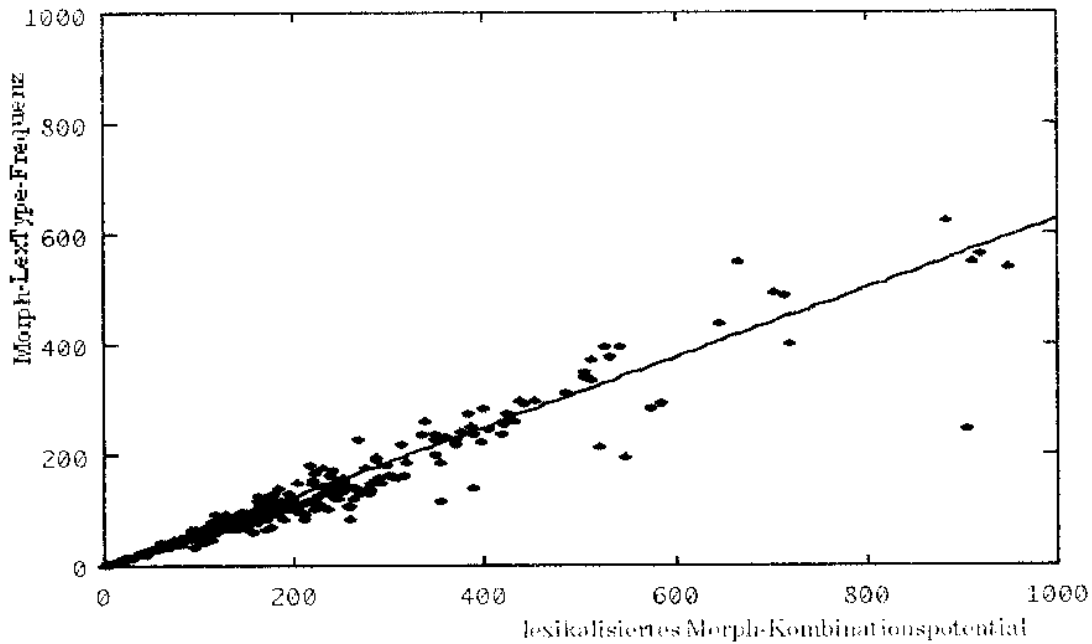


Abb. 12: Die Abhängigkeit der Morph-LexType-Frequenz vom lexikalisierten Morph-Kombinationspotential (Niederländisch) (Ausschnitt)

3.3.3 Die Abhängigkeit der Token-Frequenz von der Type-Frequenz

Die Tokenfrequenz eines Morphs bestimmt sich nach dem Modell aus seiner Type-Frequenz (vgl. Hypothese 6), der Anzahl synonyme Wörter, dem Anwendungs- und dem Invarianzbedürfnis. Die Funktionsgleichung, welche die Beziehung zwischen der Type- und der Tokenfrequenz eines Morphs m beschreibt, lautet $TOF_m = A \cdot TYF_m^B$. Dabei steht die Konstante B für die Auswirkungsstärke der Type-Frequenz und die Konstante A für die Kombination der übrigen Einflüsse auf die Tokenfrequenz.

Da, wie in Abb. 13 deutlich zu erkennen ist, die Daten wieder ungleich verteilt sind, wurden die Regressionsanalysen für alle drei Sprachen mit gewichteten Daten durchgeführt. Die Ergebnisse der Berechnungen sind in der Tabelle 6 wiedergegeben.

Tabelle 6
Anpassung der Funktion $L\text{-}TOF_m = A + B \cdot L\text{-}TYF_m$

<i>Sprache</i>	<i>A</i>	<i>B</i>	<i>R²</i>	<i>Sign. t</i>
Deutsch	72,2983	1,06139	0,93910	0,0000
Englisch	310,872	0,99930	0,89605	0,0000
Niederländisch	373,385	0,88186	0,91249	0,0000

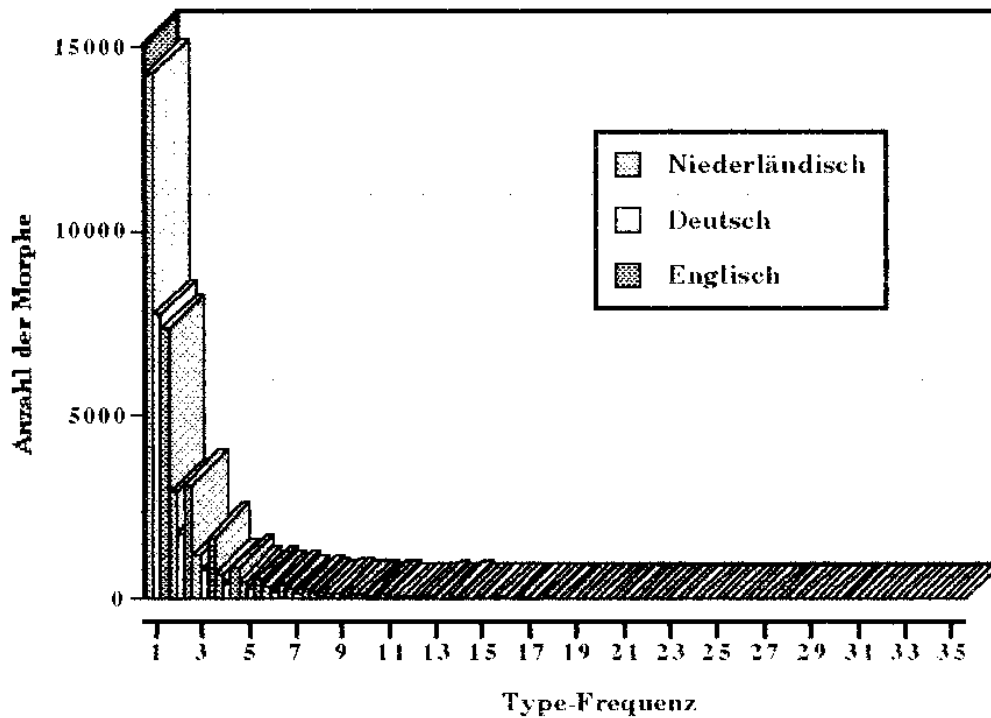


Abb. 13: Häufigkeitsverteilung der Type-Frequenz (Ausschnitt)
Die Analyseergebnisse entsprechen den Erwartungen: Der t -Test bestätigt in allen drei Sprachen die Linearität der Beziehung zwischen den Größen des linearisierten Modells, und die hohen Werte des Determinationskoeffizienten stehen für sehr gute Anpassung an die Daten, auch wenn die Güte der Anpassung leicht variiert.

Der optische Eindruck der Anpassungen, den die Abb. 14 bis 16 liefern, ist nicht so gut, wie es die Ergebnisse der Regression vermuten lassen. Die Datenpunkte liegen nur im Bereich der kleinen Morph-Type-Werte sehr nahe an der theoretischen Funktion. Je größer die Morph-Type-Werte werden, um so größer sind die Entfernungen zur Funktion. Die guten Ergebnisse der Anpassung sind auf die Gewichtung der Daten zurückzuführen. Da diese aufgrund der Häufigkeitsverteilung jedoch sinnvoll war, kann die Hypothese trotz der Streuung im Bereich der hohen Type-Frequenz-Werte angenommen werden.

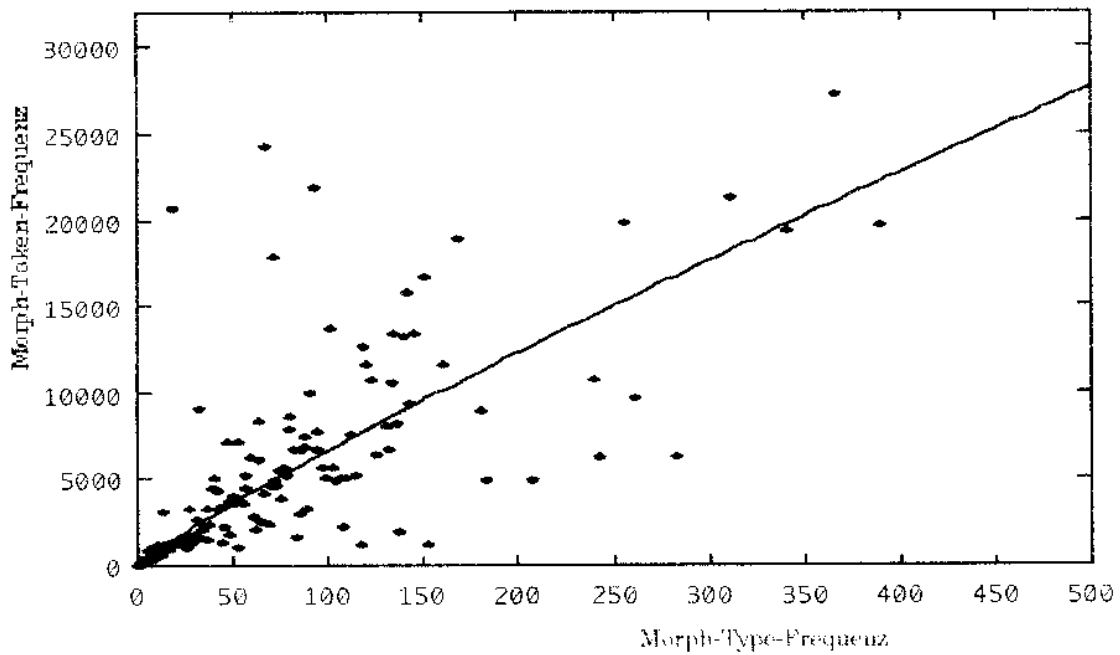


Abb. 14: Die Abhängigkeit der Token-Frequenz von der Type-Frequenz (Deutsch) (Ausschnitt)

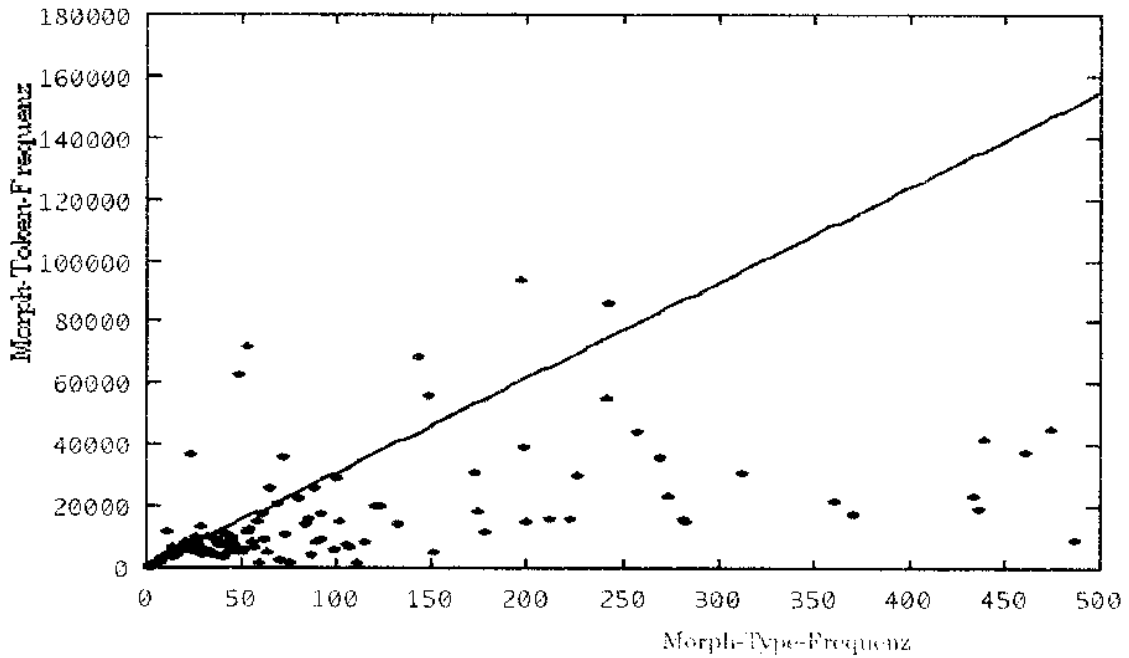


Abb. 15: Die Abhängigkeit der Token-Frequenz von der Typen-Frequenz (Englisch) (Ausschnitt)

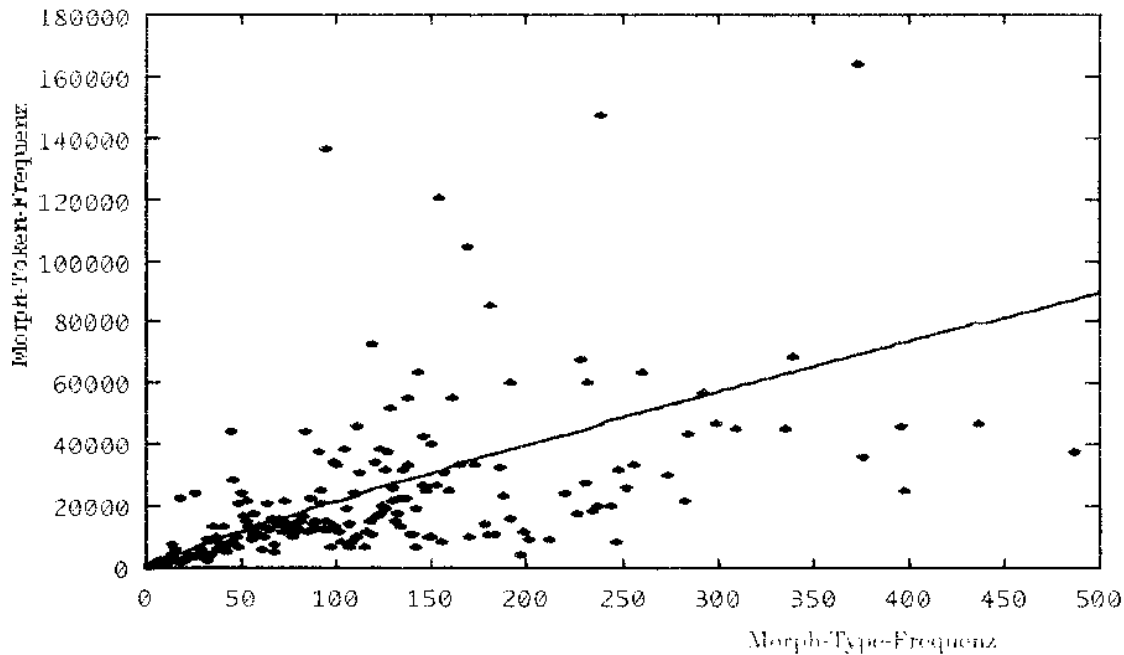


Abb. 16: Die Abhängigkeit der Token-Frequenz von der Type-Frequenz (Niederländisch) (Ausschnitt)

3.3.4 Die Abhängigkeit der Morphemlänge von der Morph-Token-Frequenz

Nach Hypothese 8 ergibt sich die Länge eines Morphs u.a. aus seiner Token-Frequenz. Formal wurde diese Hypothese durch die Gleichung $ML_m = A \cdot TOF_m^B$ ausgedrückt, wobei die Variable A dem Einfluß des Bedürfnisses $minD$ und B der Einflußstärke der Morph-Token-Frequenz auf die Länge entspricht.

Weil bei der morphologischen Analyse in CELEX einige Wörter nicht vollständig zerlegt wurden, da die Zerlegung beispielsweise als irrelevant angesehen wurde (siehe Eigennamen), wurden für die Untersuchungen dieser Arbeit einige Morpheme als solche übernommen, die noch weiter zerlegt werden könnten. Da diese ‚Morphe‘ normalerweise als semantische Einheit betrachtet werden können, wurden sie in den bisher vorgestellten Analysen auch als Morpheme behandelt. Bei den folgenden Untersuchungen wird jedoch die Morphemlänge von Interesse sein, und gerade für diese Größe liefert die bisherige Vorgehensweise verfälschte Ergebnisse. Denn die genannten ‚Morphe‘ sind, da sie noch weiter zerlegt werden können, übermäßig lang. Um die Überprüfungen der folgenden Hypothesen nicht zu beeinträchtigen, werden daher die genannten ‚Morphe‘ aus den Stichproben herausgenommen.

Für die Untersuchung der Hypothese, daß die Morphemlänge von der Morph-Token-Frequenz abhängt, wurde die Länge der Morphe in Graphemen gemessen.

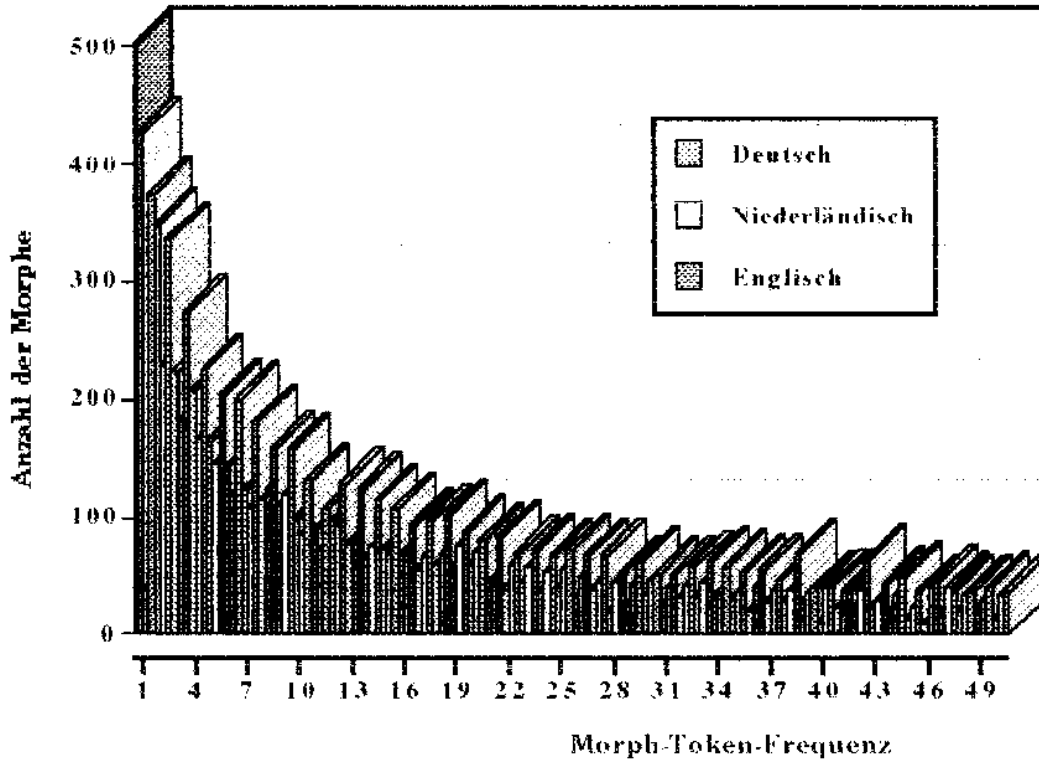


Abb. 17: Häufigkeitsverteilung der Morph-Token-Frequenz (Ausschnitt)
 Da die Daten, wie Abb. 17 in einem kleinen Ausschnitt zeigt, wiederum ungleichmäßig verteilt sind, wurde die Anpassung der Funktion mit gewichteten Daten vorgenommen. Die Ergebnisse der Anpassungen sind in Tabelle 7 wiedergegeben.

Tabelle 7
 Anpassung der Funktion $L-ML_m = \ln A + B \cdot L-TOF_m$

<i>Sprache</i>	<i>A</i>	<i>B</i>	<i>R²</i>	<i>Sign. t</i>
Deutsch	7,42350	-0,05887	0,44040	0,0000
Englisch	7,40000	-0,05120	0,41444	0,0000
Niederländisch	8,81029	-0,06993	0,45811	0,0000

Die niedrigen Werte von R^2 weisen darauf hin, daß es sich um keine gute Anpassung handelt. Aber das Signifikanzniveau des t -Tests, dessen Wert unter 0,00005 liegt, bestätigt, daß nach Linearisierung des Modells eine lineare Beziehung zwischen den Größen zu bestehen scheint. Daher sollte die aufgestellte Hypothese noch nicht abgelehnt werden. Es muß vielmehr eine Erklärung dafür

gesucht werden, warum die Daten so sehr von der theoretischen Kurve abweichen.

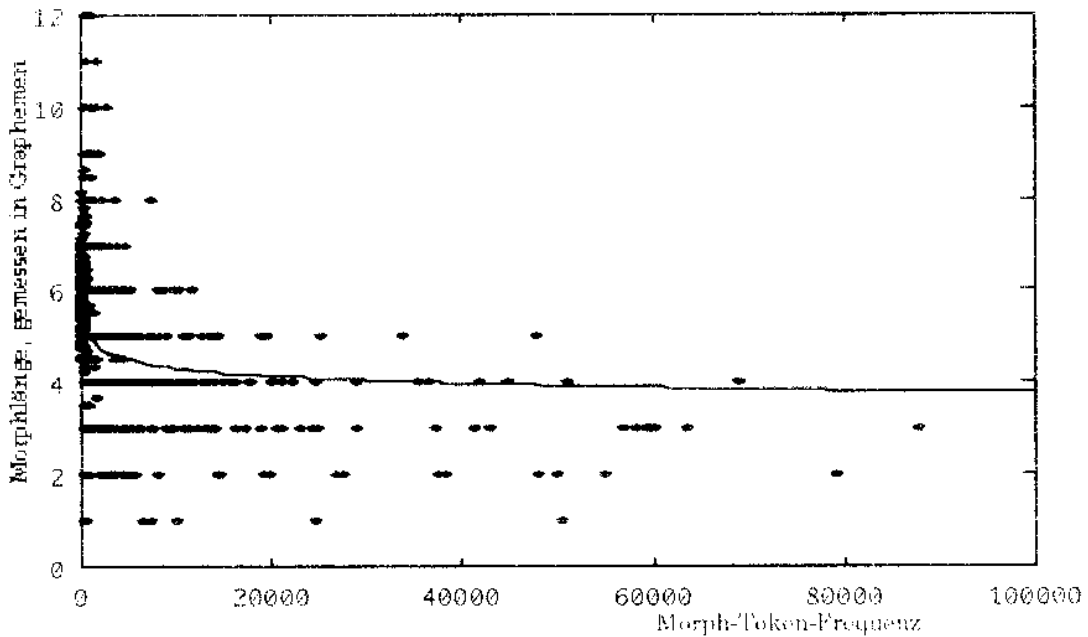


Abb. 18: Die Abhängigkeit der graphemischen Morphemlänge von der Morph-Token-Frequenz (Deutsch)

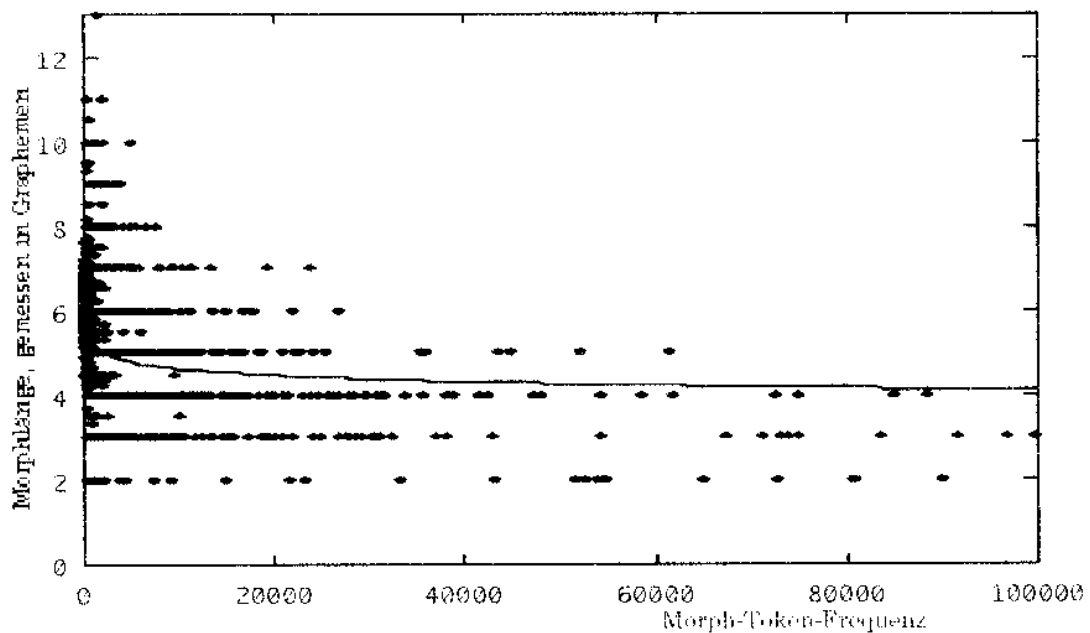


Abb. 19: Die Abhängigkeit der graphemischen Morphemlänge von der Morph-Token-Frequenz (Englisch)

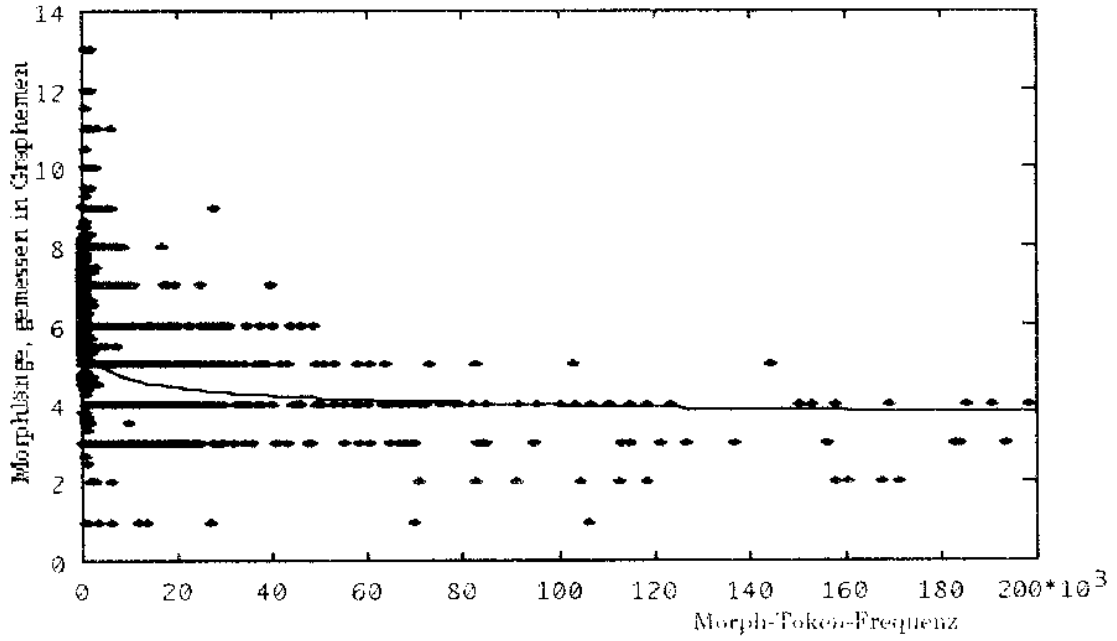


Abb. 20: Die Abhängigkeit der graphemischen Morphemlänge von der Morph-Token-Frequenz (Niederländisch)

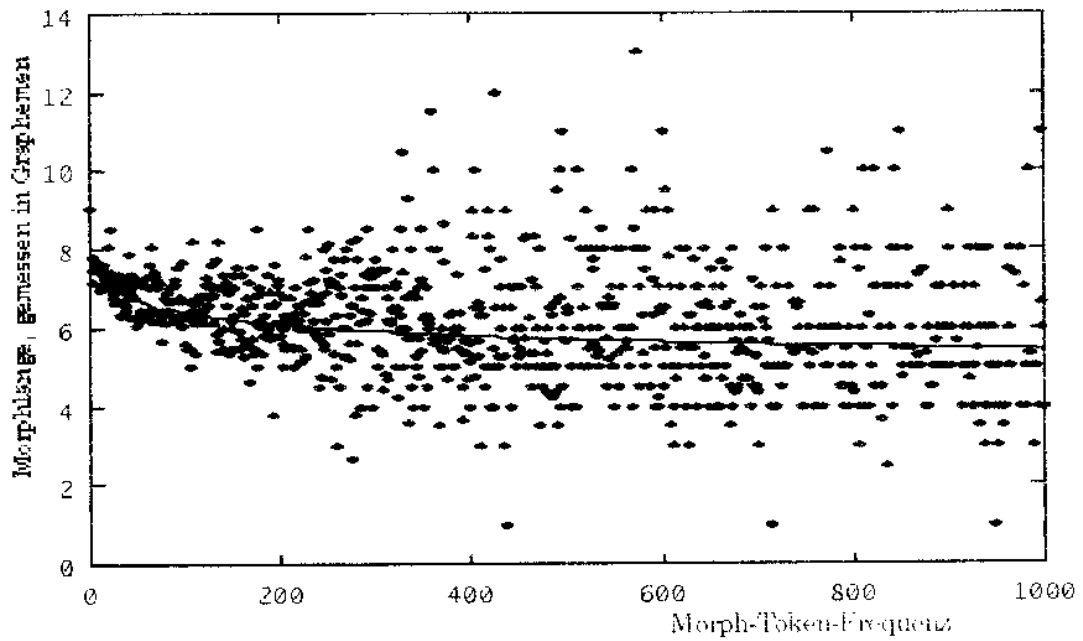


Abb. 21: Die Abhängigkeit der graphemischen Morphemlänge von der Morph-Token-Frequenz (Niederländisch) (Ausschnitt)

Die Abb. 18 bis 21 zeigen die Daten und die theoretischen Kurven für alle drei Sprachen. Dabei wurde zusätzlich ein Ausschnitt der Bereiche kleiner Frequenzwerte exemplarisch für das Niederländische herausgegriffen, um die Anpassung der Funktion an den Teil der Daten zu zeigen, der den Hauptanteil der empirisch gewonnenen Werte ausmacht.

Die Graphiken zeigen eine auffallend große Streuung der Datenpunkte, die sich bei Steigung der Token-Frequenz-Werte zunächst verstärkt, bei sehr hohen Werten jedoch wieder abnimmt. Aufgrund dieser Streuung ist es nicht verwunderlich, daß die Anpassung der theoretischen Kurve an die Daten nur mäßig gut ist. Es ist aber auch zu erkennen, daß im Bereich der kleinen Token-Frequenz-Werte, der einen sehr großen Teil der Daten ausmacht, die Datenpunkte nahe an der theoretischen Funktion liegen.

Es ist fraglich, ob die zu erkennende Streuung eine zufällige Erscheinung ist, die beispielsweise auf die dünne Besetzung des Bereichs größerer Token-Frequenzen zurückzuführen ist, oder ob die theoretische Funktion die Daten nicht ausreichend beschreibt. Es ist beispielsweise möglich, daß ein oder mehrere weitere Einflüsse für die Streuung verantwortlich sind, die im Modell noch nicht berücksichtigt wurden.

Die große Datenstreuung wurde auch von Köhler (1986) bei der Untersuchung auf der Ebene des Wortes festgestellt, u.a. bei der Abhängigkeit der Wortlänge von der Frequenz. Um zu testen, ob die Abweichungen der Datenpunkte von der theoretischen Kurve zufällig sind, wendet Köhler die Methode der gleitenden Mittelwerte an. Bei dieser Methode werden auf der Grundlage der vorhandenen (x_i, \bar{y}_i) -Paare gleitende Intervalle der Breite I gebildet, in denen auf untenstehende Weise sowohl ein mittlerer Frequenz-Wert (\bar{x}_j) als auch ein mittlerer Wert der Morphlängen (\bar{y}_j) berechnet wird.

$$\bar{x}_j = \frac{\sum_{i=j}^{j+I-1} x_i}{I} \qquad \bar{y}_j = \frac{\sum_{i=j}^{j+I-1} \bar{y}_i}{I}$$

Die so erzeugten (\bar{x}_j, \bar{y}_j) -Paare bilden eine neue Datenbasis für eine erneute Anpassung der theoretischen Kurve.

Die Abb. 22 zeigt exemplarisch in einem Ausschnitt die Datenpunkte, die sich für die niederländische Stichprobe bei einer Intervallgröße von 20 ergeben. An der Graphik ist zu erkennen, daß die Methode der gleitenden Mittelwerte die Streuung der Datenpunkte durchaus glättet. Interessanterweise nähern sich die Punkte jedoch nicht nur der Funktion an, sondern sie scheinen einer oszillierenden Kurve zu folgen. Dasselbe Phänomen stellte auch Köhler im Falle der Beziehung zwischen Wortfrequenz und Wortlänge fest. In weiteren Untersuchungen zeigt er, daß eine Vergrößerung des Intervalls eine Verstärkung des Glättungseffekts zur Folge hat. Auch dieses Phänomen kann auf der Ebene der Morphe bestätigt werden.

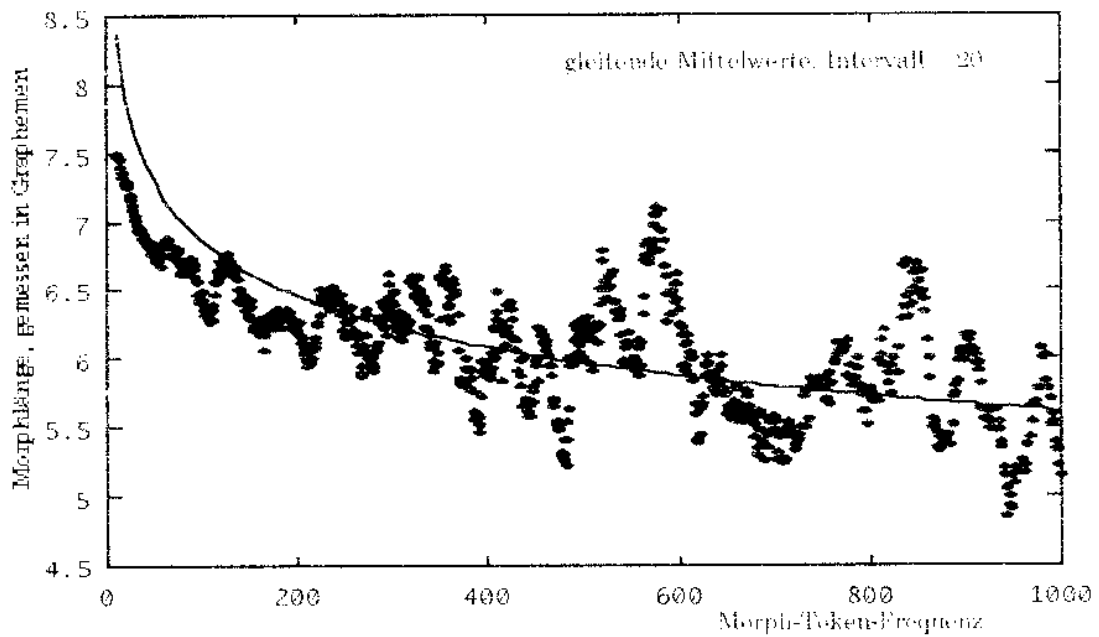


Abb. 22: Die Abhängigkeit der graphemischen Morphemlänge von der Token-Frequenz (Niederländisch) (Intervall = 20)

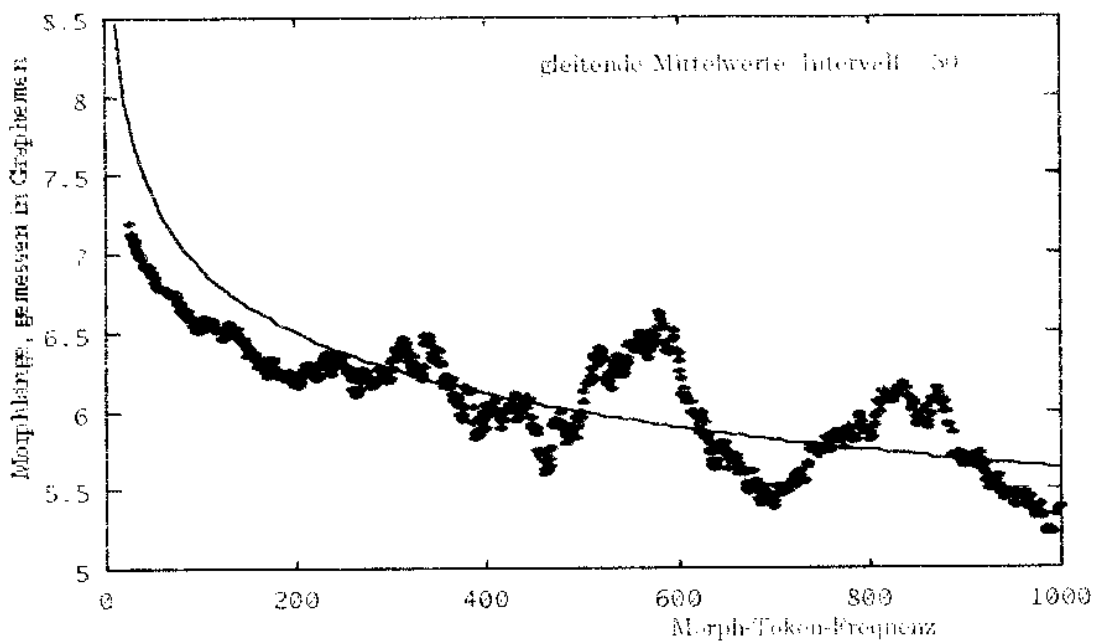


Abb. 23: Die Abhängigkeit der graphemischen Morphemlänge von der Token-Frequenz (Niederländisch) (Intervall = 50)

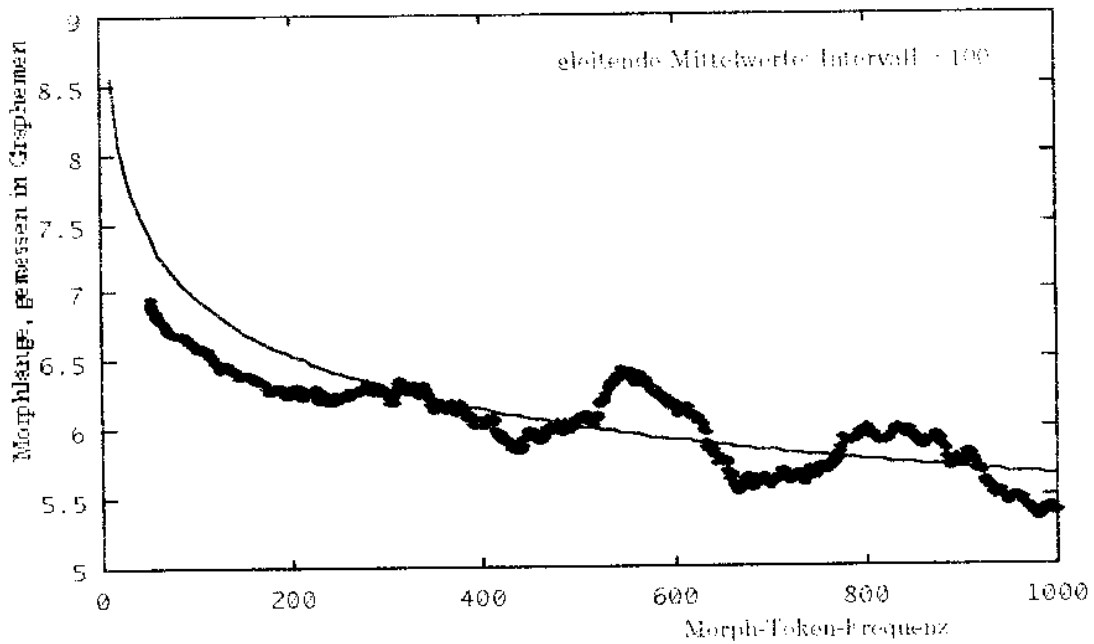


Abb. 24: Die Abhängigkeit der graphemischen Morphemlänge von der Token-Frequenz (Niederländisch) (Intervall = 100)

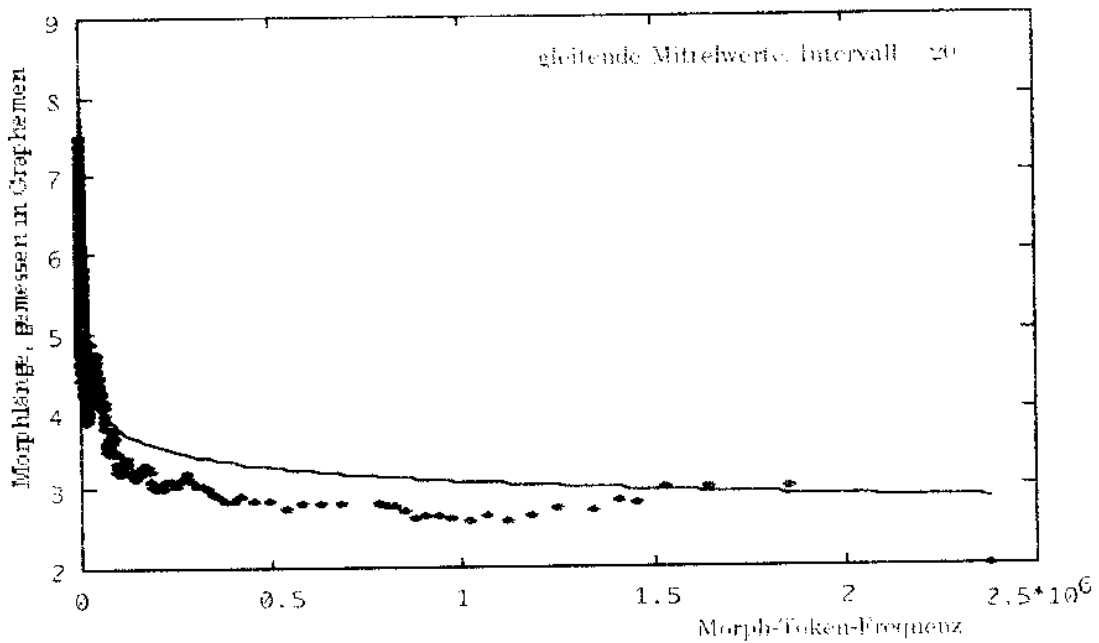


Abb. 25: Die Abhängigkeit der graphemischen Morphemlänge von der Token-Frequenz (Niederländisch) (Intervall = 20)

Die Abb. 23 und 24 sollen wiederum exemplarisch an der niederländischen Sprache den Effekt der unterschiedlichen Intervalllängen verdeutlichen. Dafür wurden Intervalle der Länge 50 und 100 ausgewählt. Damit sich die Ergebnisse besser vergleichen lassen, handelt es sich bei den Abb. 23 und 24 um denselben Ausschnitt der Token-Frequenzen wie bei Abb. 22.

Wie an den Abb. 23 und 24 zu erkennen ist, steigt mit der Größe des Intervalls auch der Wert der ersten mittleren Morph-Token-Frequenz. Das liegt ganz einfach daran, daß bei größeren Intervallen größere Frequenzwerte in die Menge der zu mittelnden Werte mit einfließen. Da die Glättung der Datenstreuung noch klarer zu erkennen ist, wenn man die gesamte Stichprobe betrachtet, zeigt die Abb. 25 alle Datenpunkte der niederländischen Stichprobe bei einer Intervallgröße von 20.

Für die beiden anderen Sprachen, Deutsch und Englisch, gelten entsprechende Ergebnisse, wie sie für das Niederländische festgestellt wurden. Es soll darauf verzichtet werden, diese im einzelnen hier darzustellen. Für einen Überblick enthält die Tabelle 8 für alle Sprachen die Ergebnisse der Anpassung nach der Anwendung der Methode der gleitenden Mittelwerte für eine Intervallbreite von 20.

Tabelle 8

Anpassung der Funktion $L\text{-}TOF_m = A + B \cdot L\text{-}TYF_m$ nach Anwendung der Methode der gleitenden Mittelwerte (Intervall = 20)

<i>Sprache</i>	<i>A</i>	<i>B</i>	<i>R²</i>	<i>Sign. t</i>
Deutsch	9,61342	-0,09427	0,83877	0,0000
Englisch	9,04055	-0,07603	0,69427	0,0000
Niederländisch	10,2381	-0,08682	0,83453	0,0000

Wie an den Werten für R^2 und an den Graphiken zu erkennen ist, handelt es sich in allen drei Fällen um eine relativ gute Anpassung, obwohl die Kurven etwas zu flach verlaufen, um an die Datenpunkte für hohe Token-Frequenz-Werte heranzureichen.

Für die ‚Schwingungen‘, die sich bei der Berechnung der gleitenden Mittelwerte einerseits bei der Untersuchung der Abhängigkeit der Morphemlänge von der Morph-Token-Frequenz und andererseits auch bei der Beziehung Wortfrequenz - Wortlänge zeigen, gibt es zwei Erklärungen: Entweder sind die Oszillationen ein inhärentes Merkmal der Daten, das auch schon in den Ursprungsdaten, d.h. vor der Bildung gleitender Mittelwerte, vorhanden, aber noch nicht so deutlich zu erkennen war, oder es handelt sich nicht um tatsächliche Schwingungen, sondern um ein Phänomen, das durch die Bildung der gleitenden Mittelwerte hervorgerufen wird.

Köhler (1986) und Zörnig et al. (1990) gehen davon aus, daß es sich tatsächlich um Oszillationen handelt. Sie liefern auch Ansätze für eine sprachwissen-

schaftliche Erklärung und mathematische Beschreibung des Phänomens: Aufgrund der Annahme, daß ein Wort um so stärker gekürzt wird, je länger es ist, beziehen sie die Kürzungsrate in die Beziehung Länge-Frequenz mit ein. Auf diese Weise wird eine oszillierende Funktion abgeleitet. Die Anpassung einer solchen Funktion durch Zörnig et al. (1990) an die Daten, die mit der Methode der gleitenden Mittelwerte erzeugt wurde, ist überzeugend. Dennoch fehlt eine ausreichende linguistische Interpretation dieser Funktion.

Es gibt auf der anderen Seite Hinweise, daß Köhlers Annahme nicht richtig ist, sondern daß es sich bei den angeblichen Schwingungen um ein Phänomen handelt, das bei der vorhandenen Art von Datenstreuung durch die Methode der gleitenden Mittelwerte erzeugt wird. Das würde bedeuten, daß die Daten keine schwingende Funktion ergeben, und es sich um eine bloße Streuung bzw. um einen noch nicht entdeckten Einfluß handelt, der nicht berücksichtigt ist. Wendet man nämlich die Methode der gleitenden Mittelwerte (Intervall = 50) auf eine Menge von Werten an, die durch zufällige Abweichung einer vorgegebenen Funktion gewonnen werden, so erhält man die Ergebnisse, die in den Abb. 26 und 27 wiedergegeben sind.

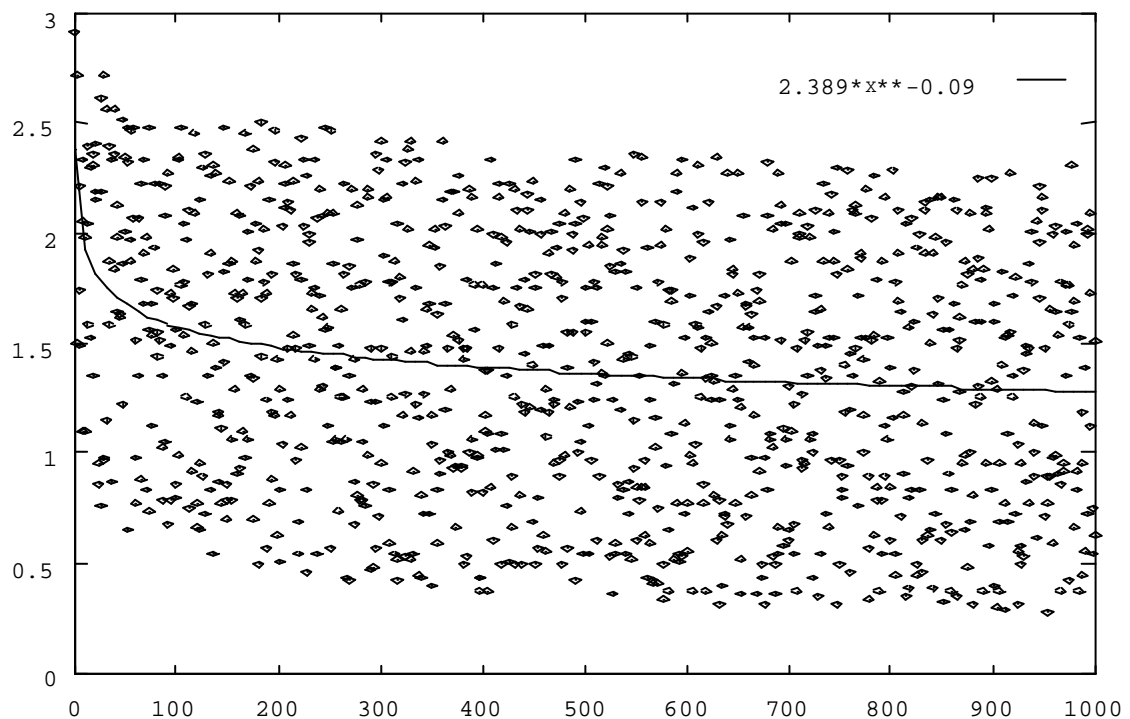


Abb. 26: Zufallsdaten im Bereich einer Beispielfunktion

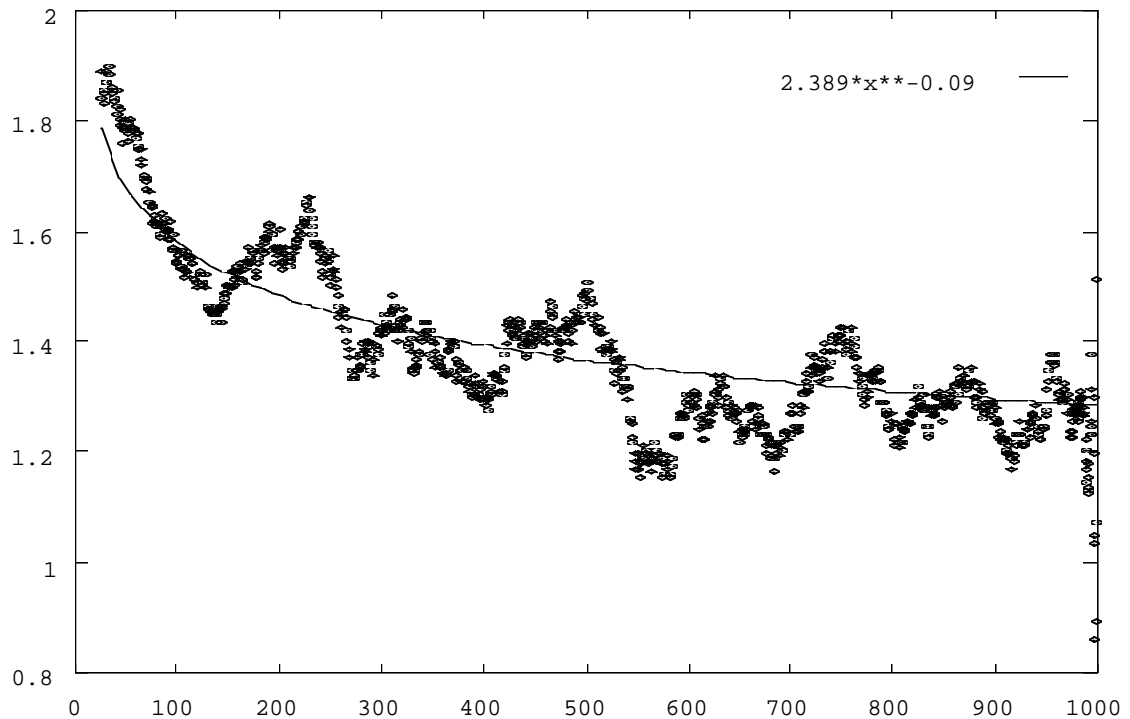


Abb. 27: Anwendung der Methode gleitender Mittelwerte (Intervall = 50) auf die Zufallsdaten

Die Abb. 27 zeigt, daß auch im Falle zufällig streuender Daten scheinbare ‚Schwingungen‘ entstehen¹². Dieses Ergebnis ist zwar kein Beweis dafür, daß die Daten der Frequenz-Längen-Beziehung keine Schwingungen enthalten, es stellt diese Annahme jedoch in Frage und unterstützt eher die Hypothese, daß es sich nicht um tatsächliche Oszillationen handelt, sondern um ein Bild, das aufgrund der immens streuenden Datenpunkte und der Methode der gleitenden Mittelwerte entsteht. Auf jeden Fall steht es im Einklang mit der Annahme, es handele sich um eine *zufällige* Streuung. Ein anderes Indiz, das gegen eine Oszillation spricht, sind die Ergebnisse einer Periodizitätsuntersuchung von Hammerl (1990:19-22). Er wendet den Phasenverteilungstest und den Phasenhäufigkeitstest von Wallis und Moore an, die eine Periodizität der empirischen Werte einer Größe in bezug auf eine andere Größe nachweisen können. Aber es konnte mit Hilfe dieser Tests keine Periodizität nachgewiesen werden.

Geht man davon aus, daß die Streuung nicht zufällig ist, so kann man sie auch anders erklären. Wie aus den Abb. 18, 19 und 20 zu erkennen ist, bleiben die Morphlängen immer unter einer bestimmten Grenze. Es ist möglich, daß gerade diese Grenze das Charakteristikum der Daten ausmacht. Das würde bedeuten, daß eine bestimmte Frequenz eine Höchstgrenze der Länge des Morphs oder

¹² Es wurde eine Reihe von Zufallsdaten getestet, um das Ergebnis zu bestätigen. Da die Ergebnisse der anderen Daten ähnlich waren, wird hier darauf verzichtet, auch diese Daten darzustellen.

des Wortes bedingt. Um diese Hypothese zu testen, müßte eine Funktion abgeleitet werden, die diese Grenze beschreibt.

Eine weitere mögliche Quelle für die große Streuung wurde noch nicht genannt, nämlich die Datenerhebung. Es wurde bei der Beschreibung des verwendeten Datenmaterials erwähnt, daß zur Bestimmung der Morph-Frequenzen grundsätzlich nur schon lexikalisierte Wörter, welche die betreffenden Morphe enthalten, berücksichtigt wurden. Weil diese Tatsache Einfluß auf die Untersuchungsergebnisse hat, ist es durchaus möglich, daß auch der Grad der Datenstreuung beeinflußt wird.

In dieser Arbeit kann nicht endgültig entschieden werden, welche der beschriebenen Hypothesen die richtige ist. Die Lösung dieses Problems liefert den Stoff für eine eigenständige Arbeit. Es sollte jedoch deutlich gemacht werden, daß die Beziehung Länge-Frequenz noch weiter analysiert werden muß.

Die in dieser Arbeit aufgestellte Hypothese über die Frequenz-Längen-Relation der Morphe konnte im Gegensatz zu der Hypothese auf der Wortebene aufgrund der Anpassungsgüte weder abgelehnt noch eindeutig bestätigt werden. Die Tatsache, daß die Anwendung der Methode der gleitenden Mittelwerte Daten erzeugt, an die sich die theoretische Funktion sehr gut anpassen läßt, deutet jedoch darauf hin, daß die aufgestellte Hypothese richtig ist, die Länge eines Morphs also bei steigender Frequenz abnimmt.

4. Zusammenfassende Betrachtung der Untersuchungen

Die empirische Überprüfung einiger Hypothesen, die im ersten Teil dieser Arbeit aufgestellt wurden, war im großen und ganzen erfolgreich, d.h. die Hypothesen wurden fast immer an einer bzw. an drei Sprachen bestätigt. Ausnahmen bilden jedoch die beiden Hypothesen 3 und 8.

Die Hypothese 3 besagt, daß das lexikalisierte Morph-Kombinations-potential eine Funktion der Morph-Polysemie ist. Eine Untersuchung aller Morphe erbrachte nur eine mäßig gute Anpassung der theoretischen Funktion an die Daten. Allerdings wurde gezeigt, daß die Aufteilung der Morphe in verschiedene Typen ganz unterschiedliche Ergebnisse lieferte. Die Einzelanalyse einiger Morphotypen (,Affixe', ,Verben', ,Nomen' und ,Adjektive') führte zu sehr guten Anpassungen, andere dagegen, deren Bedeutungsbestimmung einige Probleme aufwirft, wiesen niedrige Werte für die Anpassungsgüte auf. Als Gesamtergebnis kann daher festgehalten werden, daß zumindest für die erste Gruppe der Morphotypen die aufgestellte Hypothese bestätigt werden konnte. Hinsichtlich der zweiten Gruppe müssen noch weitere Analysen folgen. Es ist zu überlegen, ob und wie man die Erhebung der Morph-Polysemie-Daten verbessern kann bzw. darf.

Die zweite Hypothese, die nicht vollständig bestätigt werden konnte, besagt, daß die Morphemlänge eine Funktion der Morph-Token-Frequenz ist. Die Anpassung der Daten erbrachte auch in diesem Fall nur einen mittelmäßigen Wert für den Determinationskoeffizienten R^2 . Die Hypothese wurde jedoch nicht abgelehnt, da eine Anwendung der Methode der gleitenden Mittelwerte eine sehr gute Anpassung an die theoretische Funktion erbrachte. Ein Grund für die nicht sehr

gute Anpassung könnte der sein, daß kurze Morphe nicht unbedingt Bestandteile von kurzen, und damit häufigen Wörtern sind.

Alle übrigen Hypothesen wurden an allen drei Sprachen aufgrund der Ergebnisse für R^2 bestätigt.

Einige Fragen und Probleme, die sich aufgrund der Untersuchungsergebnisse stellen, konnten in dieser Arbeit jedoch nicht gelöst werden. Die wichtigste ist die folgende:

Handelt es sich bei den Untersuchungsergebnissen der Beziehung Morphlänge-Morphfrequenz um eine zufällige Streuung der Datenpunkte oder um Oszillationen, und/oder existiert ein Einfluß, der in die Betrachtungen noch nicht mit einbezogen worden ist? Das Vorhandensein von Oszillationen wurde zwar sehr in Frage gestellt, es konnte jedoch nicht vollkommen widerlegt werden.

Außer der Klärung dieser Frage stellt sich die Aufgabe, die noch nicht überprüften Hypothesen (siehe Tabelle 1) an Sprachmaterial zu testen. Außerdem ist es möglich, ähnlich wie Köhler (1986) vorgegangen ist, auch die indirekten Einflüsse, die zwischen den Größen des Modells bestehen, empirisch zu überprüfen. Für diese und für die in dieser Arbeit vorgestellten Untersuchungen sollten möglichst nicht nur die Sprachen Deutsch, Englisch und Niederländisch, sondern auch andere Sprachen in die Analyse mit einbezogen werden.

Auch wenn hier keine umfassende Überprüfung der aufgestellten Hypothesen geleistet werden konnte, versteht sich diese Arbeit als ein weiterer Schritt zur Konstruktion einer Sprachtheorie, wie sie von der linguistischen Synergetik angestrebt wird. Es wurde gezeigt, wie fruchtbar der synergetische Ansatz auch für den Bereich der Wortbildung ist, d.h. wie leicht er sich auf andere Ebenen als die von Köhler untersuchte Lexik übertragen läßt. Daher sollten weitere linguistische Bereiche auf der Grundlage dieses Ansatzes untersucht werden.

Literaturverzeichnis

- Altmann, G.** (1992). Das Problem der Datenhomogenität. In B. Rieger (Hg.), *Glottometrika 13* (S. 287-298), Bochum: Brockmeyer.
- Aronoff, M.** (1976). *Word Formation in Generative Grammar*. Cambridge, Mass.: MIT Press.
- Baayen, H.** (1992). *Markedness and Productivity: A Quantitative Analysis*. Paper presented at the Fifth International Morphology Meeting, July 1992, Krems, Austria.
- Baayen, H., Piepenbrock, R., & van Rijn, H.** (1993). *The CELEX Lexical Database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia.
- Bartsch, R., & Vennemann, T.** (1983). Grundzüge der Sprachtheorie. Eine linguistische Einführung. Unveränderte Studienausgabe d. Aufl. 1982. (Konzepte der Sprach- und Literaturwissenschaft; 32). Tübingen: Niemeyer.
- Becker, T.** (1990). *Analogie und morphologische Theorie*. München. (Studien zur theoretischen Linguistik; 11).
- Booij, G.E.** (1977). *Dutch Morphology. A Study of Word Formation in Generative Grammar*. Dordrecht: Foris.
- Bybee, J.L.** (1985). *Morphology. A study of the relation between meaning and form*. Amsterdam, Philadelphia: Benjamins. (Typological studies in language; 9).
- Derwing, B.L.** (1990). Morphology and the mental lexicon: psycholinguistic evidence. In W.U. Dressler et al. (Hg.), *Contemporary Morphology* (S. 249-265), Berlin: de Gruyter.
- Erben, J.** (1993). *Einführung in die deutsche Wortbildungslehre*. 3. Aufl. Berlin: Schmidt. (Grundlagen der Germanistik; 17).
- Fleischer, W.** (1976). *Wortbildung der deutschen Gegenwartssprache*. 4. durchgesehene Auflage. Leipzig: Bibliographisches Institut.
- Giesecking, K.** (1993). *Synergetische Aspekte von Struktur und Dynamik der englischen Lexik*. Unveröffentlichte Magisterarbeit. Universität Trier.
- Grotjahn, R.** (1992). Evaluating the adequacy of regression models: Some potential pitfalls. In B. Rieger (Hg.), *Glottometrika 13* (S. 121-172), Bochum: Brockmeyer.
- Hammerl, R.** (1990). Länge-Frequenz, Länge-Rangnummer: Überprüfung von zwei lexikalischen Modellen. In R. Hammerl (Hg.), *Glottometrika 12* (S. 1-24), Bochum: Brockmeyer.
- Henzen, W.** (1965). *Deutsche Wortbildung*. 3., durchgesehene und ergänzte Auflage. Tübingen: Niemeyer.
- Kempcke, G.** (1984). *Handwörterbuch der deutschen Gegenwartssprache*. 2. Bd. Berlin: Akademie Verlag.
- Köhler, R.** (1986). *Zur linguistischen Synergetik: Struktur und Dynamik der Lexik*. Bochum: Brockmeyer. (Quantitative linguistics; 31).

- Köhler, R.** (1989). Linguistische Analyseebenen, Hierarchisierung und Erklärung im Modell der sprachlichen Selbstregulation. In L. Hřebíček (Hg.), *Glottometrika 11* (S. 1-18), Bochum: Brockmeyer.
- Köhler, R.** (1990): Elemente der synergetischen Linguistik. In R. Hammerl (Hg.), *Glottometrika 12* (S. 179-187), Bochum: Brockmeyer.
- Krott, A.** (1999). The influence of morpheme polysemy on morpheme frequency. In *Proceedings of the Third International Conference on Quantitative Linguistics*, August 26-29, 1997, Helsinki, Finland. In: *Journal of Quantitative Linguistics* 6,1, 58-65.
- Naumann, B.** (1986). *Einführung in die Wortbildungslehre des Deutschen. 2.*, neubearb. Aufl. Tübingen: Niemeyer. (Germanistische Arbeitshefte; 4).
- Rothe, U.** (1988). Polylexy and Compounding. In K.-P. Schulz (Hg.), *Glottometrika 9* (S. 121-134), Bochum: Brockmeyer.
- Stepanowa, M.D., & Fleischer, W.** (1985). *Grundzüge der deutschen Wortbildung*. Leipzig: Bibliographisches Institut.
- van Marle, J.** (1990). Rule-creating creativity: Analogy as a synchronic morphological process. In W.U. Dressler et al. (Hg.), *Contemporary Morphology* (S. 267-273). Berlin: de Gruyter.
- Wahrig, G.** (1985). *Deutsches Wörterbuch*. München: Mosaik-Verlag.
- Wimmer, G., Köhler, R., Grotjahn, R., & Altmann, G.** (1994). Towards a theory of word length distribution. *Journal of Quantitative Linguistics*, 1.1, 98-106.
- Zörnig, P., Köhler, R., & Brinkmöller, R.** (1990). Differential equation models for the oscillation of the word length as a function of the frequency. In R. Hammerl (Hg.), *Glottometrika 12* (S. 25-40), Bochum: Brockmeyer.