

Polylexie und Kompositionsaktivität in Text und Lexik

Petra Steiner

1. Einleitung

Zu den Eigenschaften und Auswirkungen der Polylexie gibt es etliche Annahmen, die zum großen Teil in empirischen Untersuchungen bestätigt wurden. Sie betreffen Zusammenhänge mit anderen lexikalischen Eigenschaften und Häufigkeitsverteilungen. So unterliegt die Häufigkeitsverteilung der Lexeme mit einer bestimmten Polylexie bestimmten Tendenzen, deren genaues Verhalten allerdings umstritten ist (vgl. Krylov, 1982; Altmann & Schwibbe, 1989:67; Hammerl, 1991:6,139f.; Schierholz, 1991:179). In Hinblick auf andere lexikalische Eigenschaften wurden im Bereich der quantitativen Linguistik bislang die Zusammenhänge der Polylexie mit der Länge, der Frequenz und der Polytextie untersucht (vgl. Altmann, Beöthy & Best, 1982; Altmann & Schwibbe, 1989; Fickermann, Markner-Jäger & Rothe, 1984; Sambor, 1984; Sambor, 1989; Köhler, 1986). Auch zum Zusammenhang zwischen der Polylexie von Lexemen und ihrer Tendenz, Elemente von Wortbildungskonstruktionen zu sein, gibt es bereits einige Arbeiten im Bereich der quantitativen Linguistik.

Eine wichtige Grundlage für den Bereich der Komposition ist der Artikel „Hypotheses about Compounds“, in dem Altmann (1989) ausgehend von theoretischen Überlegungen elf Hypothesen aufstellt. Sie betreffen Zusammenhänge von Komposita mit der Polylexie, der Wortlänge, der Frequenz, der Polytextie und dem etymologischen Alter. Die zweite Hypothese Altmanns lautet:

„The greater the polylexy of a word the more compounds there are of which it is a component“ (Altmann, 1989:103).

Diese Hypothese wird im 4. Abschnitt dieses Beitrags vom synergetischen Ansatz ausgehend hergeleitet. Der Grund für Altmanns Annahme ist hier, daß stark polylexe Wörter häufiger verwendet werden als nicht polylexe Wörter. Dies

gilt auch für ihre Verwendung in Komposita. Diese Hypothese wurde bereits von Rothe (1988) überprüft. Als Material dienen 1858 Lexeme, die dem Wahrig Deutsches Wörterbuch (fortan: WDW) entstammen (vgl. Rothe, 1988:125). Diese Lexeme sind jeweils Bestandteil mindestens eines Kompositums im untersuchten Lexikon. Gezählt wurden dann alle weiteren Komposita mit dem jeweiligen Lexem als erster Konstituente.

Zu dieser Auswahl der Stichprobe muß angemerkt werden, daß es sich hierbei nicht um eine zufällige Stichprobe von beliebigen Lexemen handelt. Da Lexeme, die nicht Bestandteile von Komposita sind, nicht mitgezählt wurden, müssen die berechneten Werte für die durchschnittliche Anzahl von Komposita bei einer bestimmten Polylexie höher sein, als dies der Fall wäre, wenn alle Lexeme in die Untersuchung eingegangen wären.

Stimmt die zu überprüfende Hypothese, so ist vor allem der Anteil der Lexeme, die in keinem Kompositum vorkommen, bei der Menge der Lexeme mit einer geringen Polylexie in der deutschen Sprache groß. Das bedeutet, daß diese Lexeme den Schnitt der durchschnittlichen Anzahl der Komposita bei geringen Polylexiewerten beträchtlich senken. Lexeme mit einer hohen Polylexie und keinem Vorkommen in Komposita sind dagegen weniger häufig, so daß die von Rothe beobachteten Werte mit dem Anstieg der Polylexie an die Verteilung aller Lexeme angenähert werden.

Rothe (1988:125) problematisiert selbst, daß lediglich Komposita gezählt werden, die das untersuchte Lexem an erster Stelle haben. Häufig - aber nicht immer und nicht ausschließlich - dient in der deutschen Sprache die erste Konstituente zur näheren Bestimmung der zweiten Konstituente. Dieses Muster tritt häufig bei Lexemen auf, die innerhalb ihres Kontextes genauer bestimmt werden müssen. Das sind im allgemeinen Lexeme mit einer hohen Polylexie. Es muß also angenommen werden, daß der Kontrast zwischen der Anzahl der Komposita bei Lexemen mit einer hohen Polylexie und der Anzahl der Lexeme mit einer geringen Polylexie größer wäre, wenn alle Komposita berücksichtigt worden wären. In Anbetracht dieser Einschränkungen ist eine neue, entsprechend modifizierte Untersuchung des postulierten Zusammenhangs sinnvoll.¹

In der vorliegenden Arbeit wird außerdem die Auswirkung der Polylexie auf Komposita in Texten untersucht. Die Grundlage für die Hypothesen ist dabei ein Modell, wie es von Köhler (1986) auf der Basis der synergetischen Linguistik entwickelt wurde.

Vor der eigentlichen Untersuchung, die nach der „Arbeitsweise der quantitativen Linguistik“ (vgl. Altmann & Lehfeldt, 1980) strukturiert ist, werden einige relevante Arbeiten aus dem Bereich der qualitativen Linguistik dargestellt und die grundlegenden Begriffe „Polylexie“ und „Kompositum“ diskutiert.

¹ Zum Vergleich mit Rothe (1988) siehe auch Steiner (1995).

2. Qualitative Untersuchungen zur Polylexie

Auch im Bereich der qualitativen Linguistik gibt es einige interessante Untersuchungen zur Wortbildung. Schröder (1979) untersucht, welche Sememe einer lexikalischen Einheit in welchem Ausmaß für Wortbildungsstrukturen verwendet werden. Sie stellt fest, daß dies im allgemeinen vorrangig die Sememe sind, die die „Hauptbedeutungen“ tragen. Als Hauptbedeutung eines Lexems definiert sie die erste aufgeführte Bedeutung im Wörterbuch der deutschen Gegenwartssprache (Schröder, 1979:289). Dieser Teil der Untersuchung weist auf den Prozeß der Diversifikation und auf Zusammenhänge mit der Lexikologie hin.

Ferner sieht Schröder (1979:289) in der Wortbildung einerseits ein Mittel zur Reduktion von Polysemie, andererseits jedoch auch eine Quelle für die Polysemie. Dies gilt vor allem dann, wenn mehrere polyseme Lexeme Elemente einer Wortbildungsstruktur sind und sich kombinatorisch mehrere Bedeutungsmöglichkeiten ergeben.

Barz (1982) hingegen sieht diese Form von Polysemie nur bei Affigierungen gegeben, Komposita dagegen „enthalten eine größere Zahl spezieller Bedeutungsmerkmale“ (Barz, 1982:7). In ihren Arbeiten untersucht sie die *Wortbildungsaktivität* von Adjektiven des Geschmacks, der Helligkeit, der Temperatur und der Lautstärke. Als *Wortbildungsaktivität* definiert sie die Eigenschaft von Morphemen und Wörtern, als Konstituenten von WBK [d.h. Wortbildungsstrukturen, Anm. d. Verf.] fungieren zu können. [...] Ein Affix verfügt dann über einen hohen Grad der Wortbildungsaktivität, wenn es sich mit Wörtern verschiedener Wortarten sowie mit Wortgruppen verbindet und wenn es polysem ist (Barz, 1988:112).

Um die Wortbildungsaktivität des jeweiligen Adjektivs zu ermitteln, untersucht Barz sein *Wortbildungsnest*. Damit ist jeweils eine Wortmenge mit einem gemeinsamen Morphem als Konstituente gemeint, wie zum Beispiel: *kalt* - *Kälte* - *erkälten* - *Erkältung* - *Erkältungskrankheit* (vgl. Barz, 1988:96, 102). Die untersuchten Wortmengen sind verschiedenen Wörterbüchern entnommen worden, die Zahl der Sememe dem WDW (vgl. Barz, 1988:147).

Die Untersuchungen liefern drei wichtige Ergebnisse: Erstens ist der Grad der Polysemie der untersuchten Wortbildungsstrukturen tatsächlich niemals größer als der der Konstituenten (vgl. Barz, 1987:9; Barz, 1988:154f., 158). Zweitens weist Barz nach, daß bei den untersuchten Adjektiven diejenigen die größte Wortbildungsaktivität besitzen, die die meisten Sememe besitzen (vgl. Barz, 1987:9). Umgekehrt sind die Adjektive mit wenigen Sememen nicht wortbildungsaktiv: Der Grad der Mehrdeutigkeit und die Wortbildungsaktivität einer lexikalischen Einheit stehen in einem direkt proportionalen Verhältnis zueinander (Barz, 1987:11). Drittens fällt auf, daß gerade die Adjektive mit vielen Sememen häufig im Zentrum des Wortbildungsnestes stehen, wie zum Beispiel *süß* bei den Geschmacksadjektiven (vgl. Barz, 1987:9). Es gibt aber auch Fälle, in denen die

Sememzahl relativ hoch ist, es jedoch keine hohe Wortbildungsaktivität gibt (vgl. Barz, 1988:151). Barz erklärt diese Erscheinung durch die „periphere Stellung des Adjektivs“ in seinem Begriffsfeld (1988:153). Ihnen entgegengesetzt sind die „zentralen Glieder einer semantischen Klasse“, die Barz (1989:34) als solche, deren lexikalische Bedeutung die Zugehörigkeit des bezeichneten Begriffs zum Oberbegriff deutlich, eindeutig, auffällig signalisiert definiert.

Die Untersuchungen von Barz weisen darauf hin, daß der „Reflex des prototypischen Charakters“ (Barz, 1988:160) von Lexemen auf ihre Wortbildungsaktivität Einfluß hat. Barz führt allerdings kein Maß für Prototypizität ein, wie es etwa die Untersuchungen von Rosch (et al., 1976) nahelegen. Außerdem erfaßt die Arbeit nur einen sehr kleinen Teilausschnitt der deutschen Sprache, so daß weitere Untersuchungen zur Erhärtung ihrer Hypothesen notwendig sind.

3. Begriffliche Klärungen

3.1 Zum Begriff der Polylexie

Für die Mehrdeutigkeit eines Lexems gibt es in der traditionellen Sprachwissenschaft zwei Begriffe: *Polysemie* und *Homonymie* (vgl. Bußmann, 1990:593, 314; vgl. auch Duden, Grammatik, 1984:536f.; Lewandowski, 1990:818f., 405ff.). Aus diachronischer Sicht besitzen polyseme Lexeme die gleiche etymologische Wurzel, Homonyme dagegen haben verschiedene Ursprünge. Dieses Unterscheidungskriterium wird von verschiedenen Seiten kritisiert (vgl. Augst, 1975:31; Bußmann, 1990:314; Lewandowski, 1990:405ff.; Paprotté, 1983: 20f.). Aus synchronischer Perspektive ergeben sich zahlreiche Probleme bei der Zuordnung verschiedener Formen von Mehrdeutigkeit in eine der beiden Klassen². Weber (1974:18f.) weist nach, daß gängige Wörterbücher unterschiedliche Klassifizierungen besitzen. Da subjektive Entscheidungen möglichst vermieden werden sollen und zudem die Fragestellung dieser Arbeit einer Aufspaltung des Begriffs *Mehrdeutigkeit* nicht bedarf, muß eine operationale Definition gefunden werden.

Eine solche bietet Schierholz (1991:76): Er zählt die Erklärungswörter, die in der Bedeutungserklärung des Lexikons vorkommen. Im Verlauf seiner Untersuchung stößt er jedoch auf „Lemmalücken“. Das sind Erklärungswörter, die selbst nicht im Wörterbuch verzeichnet sind (Schierholz, 1991:160). Praktikabler für eine quantitative Untersuchung ist die Definition der *Polylexie* von Köhler (1986: 57, 91f.), nämlich: „die Anzahl der verschiedenen Bedeutungen, die eine lexikalische Einheit zu einem gegebenen Zeitpunkt trägt (1986:57).“ Als Schätzung für diesen Wert schlägt er „die Zahl der Wörterbucheinträge zu dieser Einheit“ (Köhler, 1986:92) vor; er bezieht sich dabei auf die Arbeit von Altmann, Beöthy und

² Eine ausführliche Darstellung zu dieser Thematik befindet sich in Schierholz (1991: 64ff.).

Best (1982). Homonyme werden dabei als ein Eintrag behandelt und die Anzahl der Einzeleinträge addiert.

Bei Verfahren, die sich auf ein bestimmtes Wörterbuch stützen, können die Untersuchungsergebnisse in Abhängigkeit vom Wörterbuch unterschiedlich ausfallen (Altmann & Schwibbe, 1989:67; Hammerl, 1991:41; Köhler, 1986:92). Schierholz (1991:178f.) schlägt deshalb vor, verschiedene Wörterbücher für die Zählungen zu verwenden. Problematisch wäre bei einem solchen Verfahren jedoch die Frage, wie die Polylexiewerte aus den verschiedenen Wörterbüchern miteinander verrechnet werden sollen. Die Bildung von mittleren Polylexiewerten auf der Basis mehrerer Wörterbücher oder die besondere Gewichtung eines bestimmten Wörterbuches müßte im jeweiligen Fall begründet werden. Zudem werden bei einer ausreichend großen Stichprobe etwaige Divergenzen in bezug auf die interessierenden quantitativen Größen wieder ausgeglichen, und für eine Untersuchung, die sprachliche Zusammenhänge aufzeigen soll, ist ein nichtvorsortiertes Wörterbuch sehr wohl geeignet.

3.2 Zum Begriff des Kompositums

3.2.1 Definition

Um einiges schwieriger als der Begriff der Polylexie ist der Begriff des Kompositums zu bestimmen. In der Wortbildungslehre ist das Kompositum neben dem Derivat, dem Präfixwort und der Zusammenbildung eine Unterklasse der *Wortbildungskonstruktionen* (vgl. Fleischer, 1983). Für den Begriff *Kompositum* als Unterklasse der Wortbildungskonstruktionen gilt als *differentia specifica*, daß es aus mindestens zwei frei vorkommenden Morphemen oder Morphemkonstruktionen zusammengesetzt ist (Bußmann, 1990:402; vgl. auch Duden, Grammatik, 1984:401; Lewandowski, 1990:578; Ludewig, 1991:115).

Derivate und Präfixwörter hingegen weisen sich dadurch aus, daß sie mittels gebundener Morpheme gebildet wurden (vgl. Bußmann, 1990:53; Duden, Grammatik, 1984:403ff.; Lewandowski, 1990:30f.; Ludewig, 1991:104f.). Häufig gilt ein Morphem dann als „gebunden“, wenn sich seine Bedeutung im Gegensatz zum freien Vorkommen des Morphems geändert hat. In der einschlägigen Literatur - einschließlich derjenigen aus dem Bereich der quantitativen Linguistik - besteht aber keine einheitliche Ansicht darüber, wann dies genau zutrifft. Deswegen wird in einigen Klassifikationen eine dritte Kategorie hinzugefügt:

Als Suffixoid oder Präfixoid werden alle Morpheme, die keiner der anderen Kategorien zugeordnet werden können, bezeichnet. Somit ist das Problem der Unterscheidung zwischen freiem Morphem und Affix verlagert auf das Problem der Unterscheidung zwischen freiem Morphem und Affixoid einerseits und auf das Problem der Unterscheidung von Affixoid und Affix andererseits (Stepanova & Fleischer, 1985:143f.), denn in der Sprachwissenschaft gibt es unter-

schiedliche Auffassungen darüber, wie die drei Klassen definiert sind und wie die Klassenzugehörigkeit eines Morphems bestimmt werden kann.

Für Fleischer (1983:67) ist die *Reihenbildung* ein Kriterium für den Übergang vom Kompositumbestandteil zum Suffix. Wie aber soll wiederum *Reihenbildung* definiert werden? Ortner und Ortner (1984:181) bezeichnen es als Reihenbildung, „wenn eine Konstituente in Bildungen nach dem gleichen Strukturmuster [innerhalb des untersuchten Korpus, Anm. d. V.] sechsmal oder öfter vorkommt.“ Fleischer faßt diesen Begriff weiter, operationalisiert ihn jedoch nicht. Vögeding (1981:106f.) gelingt eine Abgrenzung des Derivats von anderen Wortbildungskonstruktionen aufgrund phonologischer Unterschiede. Halbsuffixe und Kompositumbestandteile werden nach seinem Verfahren jedoch weiterhin durch semantische Kriterien klassifiziert, wobei diese Kriterien nicht operationalisiert werden.

Bei den beschriebenen Verfahren werden die Wortbildungskonstruktionen in Klassen eingeteilt. Allerdings ist bisweilen auch die Vorstellung einer „ordinalen Skala“ vorhanden, auf der Morpheme mehr oder weniger die „Eigenschaft Affix“ besitzen (vgl. Stepanowa & Fleischer, 1985:68). Doch auch hier gibt es keine eindeutigen Kriterien für die Einordnung auf dieser Skala. Die bisher beschriebenen Verfahren zur Einordnung von Wortbildungskonstruktionen scheitern an der Frage, wann die Bedeutung eines Morphems so stark modifiziert ist, daß von einer neuen Bedeutung gesprochen werden kann. Die Beurteilung liegt bislang letztlich im Ermessen der jeweiligen Wissenschaftlerin, und es wird immer so viele Beurteilungen wie Beurteilerinnen geben, solange es keine Operationalisierung in diesem Bereich gibt. Um den Einfluß der Subjektivität auszuschließen, wäre ein operationales Verfahren notwendig, das die Nähe der Bedeutungen der Bestandteile von Wortbildungskonstruktionen zu den Bedeutungen der Simplicia bestimmt und auf eine Ordinal- oder Intervallskala abbildet. Ein solches Verfahren gibt es aber bislang nicht.

Eine andere Möglichkeit der Unterscheidung von Komposition und Affigierung besteht darin, sie anhand ihrer Auswirkungen zu identifizieren. Als wichtige Auswirkungen der Derivation und der Präfigierung gelten die in diesem Abschnitt bereits erwähnte *Modifikation* und die *Transposition*. Die Modifikation wird als semantische Veränderung der Ausgangswörter durch Wortbildungsmorpheme definiert (vgl. Bußmann, 1990:494; Duden, 1984:403). Hauptsächlich geschieht dies durch Präfixe, aber auch durch Suffixe, wie zum Beispiel bei Diminutivbildungen durch das Morphem *-chen*.

Daß jedoch die Modifikation auch eine Auswirkung dessen ist, was allgemein unter *Komposition* verstanden wird, zeigen sowohl Ortner und Ortner (1984:69ff.), als auch Barz (1988:66ff.). Dabei kann nicht nur die erste Konstituente eines zweigliedrigen Kompositums den Inhalt der nächsten Konstituente verändern, sondern die zweite Konstituente kann auch die Bedeutung der ersten bestimmen (vgl. Ortner & Ortner, 1984:85ff.). Da die Modifikation sowohl durch

Derivation, als auch durch Komposition bewirkt werden kann, kann sie kein Unterscheidungskriterium für Komposita und Derivate beziehungsweise Präfixwörter sein. Vor allem die Derivation transponiert Ausgangswörter in eine andere Wortart (vgl. Bußmann, 1990:808) oder Wortklasse (vgl. Duden, 1984:401), wie zum Beispiel das Verb *denken* in das Adjektiv *denkbar*. Die Frage, ob auch der Prozeß der Präfigierung eine Transposition bewirkt, ist umstritten (vgl. Bußmann, 1990:598). Die Transposition scheint bei erster Betrachtung ein besserer Indikator als die Modifikation zu sein, vor allem wenn ihre Definition auf die Überführung in eine andere Wortart beschränkt ist (vgl. Bußmann, 1990:598). Nach Barz (1988:74f.) ist sie lediglich eine Auswirkung der Derivation. Der Duden (1984:475) nennt jedoch auch das Morphem *-mann*, das zur Bildung von Nomina agentis wie *Spielmann* und *Zimmermann* beiträgt, die somit Derivate seien. Die Entscheidung, ob diese Wortbildungskonstruktionen Modifikationen oder Transpositionen (oder beides?) umfassen, also Komposita oder Derivate sind, ist letztlich wieder subjektiv. Würde aber nur die Transposition in eine andere Wortart, nicht jedoch die Transposition in eine andere Wortklasse, als hinreichendes Erkennungsmerkmal für die Derivation akzeptiert werden, so entstünden Definitionslücken für Wortbildungskonstruktionen wie *rötlich* und *Schwesterschaft*, denn sie könnten keiner Klasse zugeordnet werden. Daher ist die Abgrenzung der Komposition zur Derivation anhand der Erscheinung der Transposition nicht möglich.

Es konnte gezeigt werden, daß weder die subjektive Einschätzung der semantischen Veränderungen von Morphemen in Wortbildungskonstruktionen noch die Identifikation von Wortbildungsprozessen anhand ihrer Auswirkungen zu einer ausreichenden Definition des Begriffs *Kompositum* führen. Deshalb soll im folgenden ein Verfahren diskutiert werden, das den semantischen Bereich nicht berücksichtigt. Einen solchen Weg schlägt Rothe vor:

„In order to distinguish a component from affixes, prefixes or further morphological elements, we took as criterion its autonomous existence in the German lexicon, that is, its existence as a lexeme that can also be used in isolation.“ (1988:125).

Mit anderen Worten: (Definition:) Eine Wortbildungskonstruktion ist genau dann ein Kompositum, wenn für ihre Komponenten Lexeme im Lexikon verzeichnet sind.

Die konsequente Anwendung dieser Definition hat zur Folge, daß auch Verben mit trennbaren Partikeln, wie *ab-* und *an-*, als Komposita klassifiziert werden. Zu diesem Punkt gibt es gegensätzliche Auffassungen: Fleischer bezeichnet „Bildungen, die gewöhnlich zu den „Partikelkomposita“ gerechnet werden“ (Fleischer, 1983:333) als Präfixwörter, mit Hinweis auf die Veränderung ihrer Bedeutung innerhalb der Wortbildungskonstruktionen.

Im Duden (1984) gibt es zu diesem Thema Widersprüche: Bestimmte Wortbildungskonstruktionen fallen sowohl unter die Kategorie ‚Verbzusammensetzung‘ als auch unter die Kategorie ‚Zusatz von Präfixen und Halbpräfixen‘ (Duden, 1984:414f., 426ff.).

Von Polenz (1980:173) macht auf die Unterschiede zwischen „echten Präfixverben“, wie *erblühen*, und „Partikelverben“, wie *aufblühen*, auf-merksam. Augst (1975:54) faßt trennbare Partikel als Zusammensetzungen auf. Ludewig (1991:116) rechnet Wortbildungskonstruktionen aus Partikel und Verb zu den Zusammensetzungen, ebenso wie Shaw (1979:60; 171f.). Die obige Definition führt also zur gleichen Klassifizierung der Partikelverben, wie sie auch in einem Teil der Literatur vorgeschlagen wird.

Ein wichtiger Punkt ist die Frage, wie Morpheme wie *-ei* (zum Beispiel in *Kumpanei*), *-schaft* (zum Beispiel in *Burschenschaft*) und *-haft* (zum Beispiel in *lachhaft*) behandelt werden sollen. Es handelt sich um gebundene Morpheme, die homonym zu freien Morphemen sind; von einer mehr oder weniger großen semantischen Modifikation durch die Wortbildung kann hier nicht gesprochen werden. In diesen Fällen muß bei jeder Wortbildungskonstruktion entschieden werden, ob es sich um freie oder gebundene Morpheme handelt.

Abgesehen von diesen wenigen Ausnahmen ist die operationale Definition des Begriffs *freies Morphem* als ‚autonom vorkommende Einheit im Lexikon‘ im Vergleich mit anderen Ansätzen ein handhabbares Mittel für eine Untersuchung im Bereich der quantitativen Linguistik. Mit ihrer Hilfe können Wortbildungskonstruktionen den Klassen *Kompositum* oder *Affigierung* zugeordnet werden.

3.2.2 Zur Motiviertheit und Paraphrasierung von Komposita

Als eine wichtige Eigenschaft einer Wortbildungskonstruktion wird häufig der Grad ihrer *Motiviertheit* genannt. Rothe (1988:123) differenziert zwischen motivierten, teilmotivierten und nichtmotivierten Komposita. Als Unterscheidungskriterium dient die *Paraphrasierungsmethode*. Im folgenden soll die Relevanz des Begriffs ‚Motiviertheit‘ und der Paraphrasierungsmethode für die Untersuchung geklärt werden.

Eine Wortbildung [gemeint ist: eine Wortbildungskonstruktion, Anm. d. V.] gilt als motiviert, wenn sich ihre Gesamtbedeutung aus der Summe der Bedeutungen ihrer einzelnen Elemente ableiten läßt (Bußmann, 1990:507). Motiviertheit - auch *Motivation* (Fleischer, 1983:11ff.) - gilt als ein Merkmal, mit dessen Hilfe aus synchroner Sicht Simplicia von Komposita unterschieden werden können. In zahlreichen Ansätzen wird jedoch davon ausgegangen, daß Motivation schlechthin *das* Kennzeichen von Komposita ist: Ein Wort ist ein Kompositum, wenn in der Wortbildungsparaphrase beide Glieder getrennt auftreten (Sambor, 1989:184).

„The most salient property of word-formations is their morphologically analysable and motivated character: we have to deal with syntagmas which are based on a determinant/de-terminatum (modifier/head) relationship at the most abstract and general level of semantic interpretation [...], and whose meanings ideally can be derived from their form, more precisely from the meanings of their constituents and the underlying morphosemantic patterns. In this respect, word-formation syntagmas differ crucially from monomorphemic lexical items, which are not morphosemantically transparent.“ (Kastovsky, 1990:76).

Jedoch schränkt Kastovsky selbst ein, daß es „no neat and clear-cut dividing line between opaque and transparent lexical items“ (1990:76) gibt.

Die Annahme, die Bedeutung eines Kompositums müsse sich aus seinen Bestandteilen ermitteln lassen (oder es handle sich eben um kein Kompositum), hängt mit dem strukturalistischen Ansatz in der Wortbildung zusammen, nach dem Komposita „irgendwie von Phrasen oder Sätzen abgeleitet“ (Shaw, 1979: 39) sind.

Die Methode, um die Motiviertheit eines Kompositums zu „ermitteln“, ist im allgemeinen seine Paraphrasierung (vgl. Duden, 1984:389). Ortner (1984:151ff.) bemerkt zu diesem Verfahren jedoch kritisch, daß es in der bisherigen Literatur zur Wortbildung keine Konventionen dafür gebe, was die jeweils richtige Paraphrase sei. So gibt es verschiedene Paraphrasen für gleiche Wortbildungskonstruktionen, und Ortner (1984:153) folgert daraus, daß es „*die* Paraphrase nicht gibt [Herv. d. Autors]“. Entsprechendes äußert auch Heringer (1984). Besonders problematisch ist die Paraphrasierung von Wortbildungskonstruktionen wie *Spielanzug*. Dieses Beispiel ist „morphologisch doppelt motiviert“ (Ortner & Ortner, 1984:118), denn die erste Konstituente kann sowohl auf einen Verbalstamm als auch auf ein Substantiv zurückgeführt werden (vgl. auch Augst, 1975: 61; Duden, 1984:447).

Von anderer Seite wird kritisiert, daß die Paraphrasierungsmethode auf sogenannte *Determinativkomposita*, bei denen die erste Konstituente die zweite begrifflich eingrenzt, zugeschnitten ist (vgl. Herbermann, 1981:183; Höhle, 1985: 325f.). Auch sogenannte *Kompositionsmetaphern*, die häufig in der Lyrik und in Zeitungstexten vorkommen, sind nur sehr bedingt paraphrasierbar (vgl. De Knop, 1987; Ortner & Ortner, 1984:163f.).

Zu all diesen Kritikpunkten kommt hinzu, daß es bezüglich der Motivation unterschiedliche Stufen und Zwischenstufen gibt (vgl. Shaw, 1979). Ähnlich wie bei der Polarität Komposition - Affigierung werden alle Wortbildungskonstruktionen, die weder motiviert noch unmotiviert sind, mit einer dritten Kategorie, nämlich als *teilmotiviert*, bezeichnet. Und ähnlich wie bei der bereits diskutierten Aufteilung bereiten die Zuordnungen zu diesen drei Klassen Schwierigkeiten (vgl. Vögeding, 1981:22). Auch Einteilungen in fünf verschiedene Motivationsstufen (vgl. Luukkainen, 1990:304) sind keine Verbesserung: So nennt Luukkainen die Wortbildungskonstruktion *Augenblick* als Beispiel für die zweite, noch relativ

durchsichtige Motivationsstufe, während Fleischer (1983:14) diese als ein semantisch unmotiviertes Kompositum einstuft:

Eine idiomatisierte [...] Zusammensetzung wie etwa Augenblick läßt immerhin in ihrer Struktur noch die beiden Wörter Auge und Blick erkennen, ist also im Vergleich zu einem Wort wie Moment wenigstens formal, äußerlich strukturell motiviert.

Und Stepanowa (Stepanowa & Fleischer, 1985:162) bezeichnet Bildungen wie *Junggeselle* und *Frauenzimmer* als „Komposita [!] mit völlig oder stark verdunkelter Motiviertheit“.

Aufgrund der dargelegten Kritikpunkte wird die Motiviertheit nicht weiter in die Untersuchung einbezogen.

4. Die linguistischen Hypothesen

4.1 Die Polylexie und die Kompositionsaktivität

Der Zusammenhang zwischen Polylexie und Kompositionsaktivität sowie der zwischen Polylexie und der Anzahl der Komposita mit einem bestimmten Lexem im Text sollen in diesem und im nächsten Abschnitt auf der Grundlage der synergetischen Linguistik abgeleitet werden. Ausgangspunkt ist eine bereits bestehende Hypothese zum Zusammenhang zwischen Polylexie und Phrasenlänge (vgl. Köhler, 1990a). Das Verhältnis zwischen diesen beiden sprachlichen Eigenschaften wird durch das Spezifikationsbedürfnis beeinflusst. Es repräsentiert „die Notwendigkeit, gegebene Mehrdeutigkeiten und Vagheiten einer lexikalischen Einheit zu verringern“ (Köhler, 1990a:3).

Es gibt mehrere Möglichkeiten, wie Mehrdeutigkeiten verringert werden können (vgl. Köhler, 1986:60). Eine Möglichkeit besteht darin, ein neues Wort für den Begriff zu (er-)finden (lexikalische Methode). Es ist aber auch möglich, aus dem bereits bestehenden lexikalischen Inventar eine Wortbildungskonstruktion zu schaffen (morphologisches Mittel). Ferner kann eine nähere Bestimmung mit Hilfe eines syntaktischen Gefüges geschehen, zum Beispiel durch einen Relativsatz. In der deutschen Sprache selten ist die vierte Möglichkeit der Spezifikation mittels prosodischer Mittel. Köhler (1986:60) versteht diese vier Möglichkeiten als funktionale Äquivalente, die das Spezifikationsbedürfnis erfüllen können (vgl. Köhler, 1987b:247; 1990b:98).

Die Möglichkeit der Bedienung des Spezifikationsbedürfnisses mit syntaktischen Mitteln wirkt nach Köhler (1990a:10f.) auf den Zusammenhang zwischen Polylexie und Phrasenlänge. Denn je größer eine Mehrdeutigkeit ist, desto notwendiger ist es, sie zu verringern. Dies kann beispielsweise mittels Appositionen, Relativsätzen, Genitivattributen oder Personalpronomina geschehen:

Die Längenzunahme einer Phrase ist proportional zur Polylexie der Wörter in dieser Phrase (Köhler, 1990a:10).³ Köhler leitet eine Gleichung der Form

$$(1) \quad PhL = aPl^b \quad (PhL: \text{Phrasenlänge})$$

ab, wobei es sich bei b um das Ausmaß handelt, mit dem in der jeweiligen Sprache das Spezifikationsbedürfnis mit syntaktischen Mitteln bedient wird. Diese Eigenschaft wird im weiteren als *Analytizität* bezeichnet (vgl. Abbildung 1).

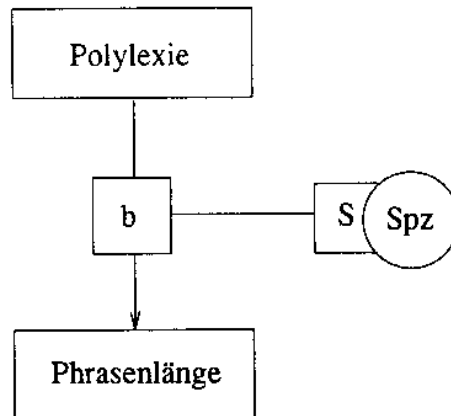


Abb. 1: Der Zusammenhang zwischen Polylexie, Phrasenlänge und dem Spezifikationsbedürfnis (nach Köhler, 1990a:12).

Zwischen der Polylexie eines Lexems und seiner Wortbildungsaktivität (vgl. Abschnitt 2) besteht ein Zusammenhang, der analog zu dem zwischen der Polylexie und der Phrasenlänge ist: (Hypothese:) Je größer die Polylexie eines Lexems ist, desto größer ist seine Wortbildungsaktivität.

Diesen Zusammenhang hat im wesentlichen schon Barz (1987:11; s.o.) postuliert. Für Komposita als Teilmenge der Wortbildungsstrukturen wird zur obigen Hypothese ein Spezialfall formuliert: (Hypothese:) Je größer die Polylexie eines Lexems ist, desto größer ist seine Kompositionsaktivität (*K*-Aktivität). Diese Hypothese stimmt mit der zweiten Altmanschen Hypothese zur Polylexie und Kompositionsaktivität überein (s.o.).

„Compounding is a specification, a narrowing of the (extension of the) meaning“, schreibt Altmann (1989:103).

Deshalb haben die Komposita im Durchschnitt auch weniger Bedeutungen als ihre Komponenten. In Abbildung 2 wird der Zusammenhang zwischen Polylexie und Kompositionsaktivität dargestellt und in Verbindung mit dem Zusammen-

³ Eine Einschränkung muß bei Relativsätzen gelten, weil diese Lexeme spezifizieren, die nicht selbst in der Relativphrase enthalten sind.

hang zwischen Polylexie und Phrasenlänge gebracht. Dabei bezeichnet b die Analytizität und k den Anteil der Komposition an der Synthetizität der Sprache.

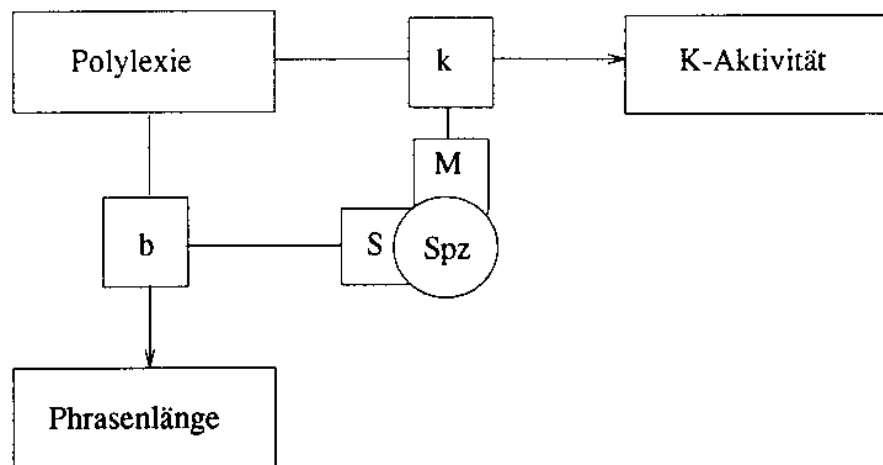


Abb. 2: Die Auswirkungen der Polylexie in Verbindung mit zwei funktionalen Äquivalenten des Spezifikationsbedürfnisses.

Nach Köhler (1990a:16) schließen sich die verschiedenen Mittel zur Bedienung des Spezifikationsbedürfnisses nicht aus. Deshalb kann im Prinzip ein und dasselbe Lexem sowohl mit morphologischen als auch mit syntaktischen Mitteln spezifiziert werden. Bisweilen besteht eine Konkurrenz zwischen einer Wortgruppe und einer Wortbildungskonstruktion wie zum Beispiel bei *leicht husten - hüsteln* und *großer Teil - Großteil* (Beispiele aus Stepanowa & Fleischer, 1985: 189). Nach Altmann (1989:100) erfüllt die Bildung eines Kompositums im Vergleich etwa zur Bildung eines Relativsatzes das Bedürfnis nach Minimierung des Produktionsaufwandes (MinP) (vgl. Köhler, 1986:20, 76), das auf die Länge wirkt, in diesem Fall auf die Satzlänge. Wird ein Kompositum gebildet, so wird dadurch die Satzlänge gekürzt, denn der jeweilige Begriff muß nun nicht mehr als Präpositional- oder Relativphrase spezifiziert werden. Andererseits bedingt die Verringerung des Produktionsaufwandes auf der syntaktischen Ebene eine Erhöhung des Produktionsaufwandes auf der morphologischen Ebene, denn die Anzahl der Wörter wird zwar geringer, die Zahl ihrer Silben aber größer. Da das Bedürfnis nach Minimierung des Produktionsaufwandes mehrere komplementäre Ebenen betrifft, kann es für eine Ebene nur auf Kosten der anderen Ebenen bedient werden. Für den Zusammenhang zwischen Polylexie und Phrasenlänge beziehungsweise Kompositionsaktivität bedeutet das, daß die Auswahl des spezifizierenden Mittels beeinflußt werden kann. Zum Beispiel kann das Bedürfnis nach Minimierung des Produktionsaufwandes auf morphologischer Ebene stärker sein als das auf der syntaktischen Ebene. In diesem Fall würde kein Kompositum zur Spezifikation eines Lexems gewählt, sondern eher eine Phrase.

Ferner konkurrieren die Systembedürfnisse auf der gleichen Ebene: Beispielsweise kann das Bedürfnis nach Minimierung des Produktionsaufwandes auf der morphologischen Ebene geringer ausgeprägt sein als das Spezifikationsbedürfnis, so daß zahlreiche Wortbildungskonstruktionen entstehen, obwohl dies nicht ökonomisch ist. Dieser Zusammenhang wird in Abbildung 3 dargestellt, im Rahmen dieses Beitrags jedoch nicht untersucht.

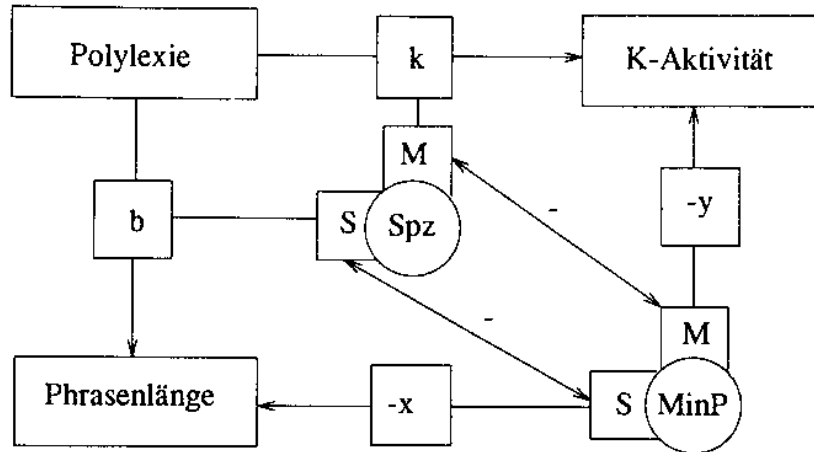


Abb. 3: Zusammenhang zwischen dem Spezifikationsbedürfnis und dem Bedürfnis nach Minimierung des Produktionsaufwandes und ihre Auswirkungen auf die Phrasenlänge und die Kompositionsaktivität.

4.2 Die Polylexie und die Anzahl der Komposita im Text (TextLK-Aktivität)

Der Zusammenhang zwischen Polylexie und Kompositionsaktivität wirkt sich auch auf den Textbereich aus: Je größer die Kompositionsaktivität eines Lexems ist, desto stärker ist die Tendenz, daß es in einem Text als Konstituente eines Kompositums oder mehrerer Komposita auftritt. Wie oft das Lexem in verschiedenen Komposita in einem Text vorkommt, hängt vom Anwendungsbedürfnis (Anw) dieses Lexems ab. Das Anwendungsbedürfnis repräsentiert die für jede lexikalische Einheit unterschiedliche Verwendungstendenz durch den Sprecher, die hauptsächlich von der Relevanz des jeweiligen Designats im gegebenen Kontext abhängt (Köhler, 1987a:162; vgl. auch Köhler, 1986:67). Das Anwendungsbedürfnis bezieht sich in diesem Fall nur auf das Lexem, nicht aber auf die Komposita, in denen es erscheint. Abbildung 4 veranschaulicht den indirekten Zusammenhang zwischen der Polylexie und der TextLK-Aktivität.

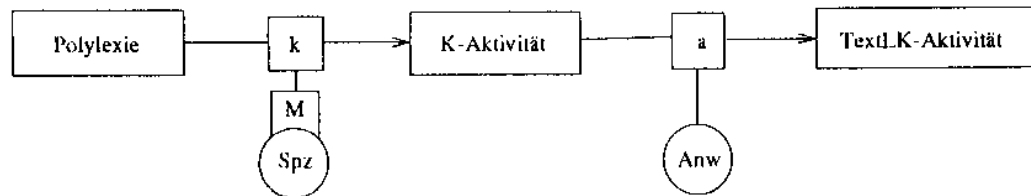


Abb. 4: Der indirekte Zusammenhang zwischen der Polylexie und der TextLK-Aktivität.

4.3 Weitere Eigenschaften

4.3.1 Textspezifische Eigenschaften

Sowohl die Teilsatz- und Phrasenlänge als auch die TextLK-Aktivität hängen von der jeweiligen Textsorte ab. Pieper (1979:65ff.) stellt anhand von Untersuchungen an 160 Texten fest, daß Hörspiele eine durchschnittliche Clauselänge von 5,88 Wörtern, allgemeine Gesetzestexte jedoch eine durchschnittliche Clauselänge von 13,01 Wörtern besitzen. Nach Altmann (Altmann & Schwibbe, 1989:43) ist bei der Beziehung zwischen Teilsatz- und Satzlänge

$$(2) \quad TsL = aSl^b b < 0 \quad (TsL: \text{Teilsatzlänge}, Sl: \text{Satzlänge})$$

der Parameter a „ein Parameter des Stils, des Genre, des Autors usw.“ Auch das Vorkommen von Komposita ist textsortenspezifisch. Beispielsweise besitzen fachsprachliche Texte mehr Komposita als andere Texte (vgl. De Cortes, 1982; Ortner & Ortner, 1984:183). Je nach Textsorte kann die Funktion eines Kompositums innerhalb des Textes wichtig sein. Zum Beispiel können Komposita die Redundanz des Textes erhöhen oder Textzusammenhänge verdeutlichen (vgl. Dederding, 1982, 1983; Ortner & Ortner, 1984:193; Schonebohm, 1979; Schröder, 1978; Wildgen, 1980).

4.3.2 Sprachspezifische Eigenschaften

Die Phrasenlänge und die Kompositionsaktivität werden nicht nur vom Spezifikationsbedürfnis und dem Bedürfnis nach Minimierung des Produktionsaufwandes bestimmt. Vielmehr wirkt in jeder Sprache auch das Kodierungsbedürfnis (vgl. Köhler, 1990b:104; 1991:47). Nach Köhler umfaßt das Kodierungsbedürfnis die Bühlerschen „Funktionen“ der Sprache, nämlich Ausdruck, Appell und Darstellung. Ohne dieses grundlegende Bedürfnis würde es keine Sprache geben. Wie das Spezifikationsbedürfnis (vgl. 4.1) kann auch das Kodierungsbedürfnis durch mehrere funktionale Äquivalente befriedigt werden. Hier werden lediglich die k-

xikalischen, die morphologischen und die syntaktischen Mittel betrachtet; die prosodischen Mittel interessieren in diesem Zusammenhang nicht.

Lexikalische Mittel zur Bedienung des Kodierungsbedürfnisses können beispielsweise die Lexikalisierung von Wortbildungskonstruktionen und von Neologismen sein. Andere Möglichkeiten sind die Entlehnung und die Erhöhung der Polylexie bereits existierender Wörter (vgl. Köhler, 1990b:105f.; 1991:47f.). Alle diese Methoden wirken sich auf die Lexikongröße aus: Bis auf die Erhöhung der Polylexie vergrößern sie das Lexikon. Ob der Einfluß der lexikalischen Mittel auf die Lexikongröße proportional oder umgekehrt proportional wirkt, hängt also vom Anteil der Erhöhung der Polylexie im Vergleich zu den anderen Methoden ab.

Morphologische Mittel zur Bedienung des Kodierungsbedürfnisses sind beispielsweise die Komposition und die Derivation. Je größer der Anteil der morphologischen Mittel an der Bedienung des Kodierungsbedürfnisses ist, desto größer ist auch die Wahrscheinlichkeit, daß ein Lexem Element von Wortbildungskonstruktionen wird - also seine Wortbildungsaktivität. Die Kompositionsaktivität hängt davon ab, welche Art von Wortbildungskonstruktion in der jeweiligen Sprache bevorzugt verwendet wird. Köhler (1990b:107; 1991:49) unterscheidet hier zwischen synthetischen und analytischen Kodierungsmethoden. Die analytischen Methoden haben deskriptivere Wortbildungskonstruktionen zur Folge als die synthetischen, die generierten Wörter sind jedoch länger.

Die Verteilung der Kompositionsaktivität der Wörter einer Sprache hängt also davon ab, in welchem Maße das Kodierungsbedürfnis mit morphologischen, analytischen Methoden bedient wird. Analytische und synthetische Kodierungsmethoden müssen sich dabei nicht ausschließen. In der deutschen Sprache gibt es beispielsweise eine Konkurrenz zwischen den Morphemen *-er* und *-mann*, wie zum Beispiel bei *Säer* und *Sämann* (vgl. Weber, 1990:136f.).

Syntaktische Mittel zur Bedienung des Kodierungsbedürfnisses sind Parataxe und Hypotaxe (vgl. Köhler, 1991:48f.). Auch hier unterscheidet Köhler zwischen analytischen und synthetischen Kodierungsmethoden (vgl. auch Köhler, 1990b:107f.). Während die analytischen Methoden die durchschnittliche Phrasenlänge vergrößern, bewirken die synthetischen Methoden eine Kürzung der durchschnittlichen Phrasenlänge. Die Verteilung der Phrasenlänge ist also abhängig von dem Maß, mit dem das Kodierungsbedürfnis von syntaktischen Mitteln bedient wird, sowie dem Anteil analytischer und synthetischer Methoden dieser syntaktischen Mittel. Abbildung 5 veranschaulicht die bislang beschriebenen Zusammenhänge zwischen dem Kodierungsbedürfnis und den Verteilungen der Kompositionsaktivität und der Phrasenlänge sowie der Lexikongröße.

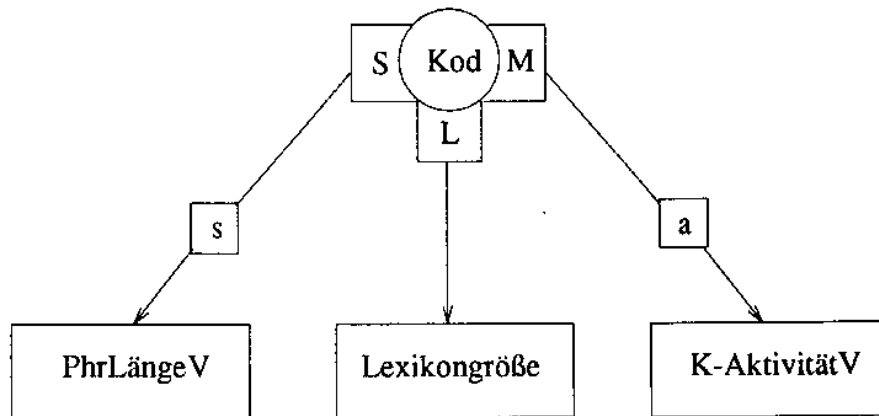


Abb. 5: Auswirkungen des Kodierungsbedürfnisses mit drei funktionalen Äquivalenten auf die Lexikongröße, die Verteilung der Kompositionsaktivität und die Verteilung der Phrasenlänge.

- a*: Anteil der analytischen Methoden bei den morphologischen Mitteln in einer Sprache;
- s*: Anteil der synthetischen Methoden bei den syntaktischen Mitteln in einer Sprache.

Es sei außerdem erwähnt, daß die Verteilung der verschiedenen funktionalen Äquivalente und Kodierungsmethoden vom Stil eines bestimmten Zeitraums abhängen kann. So können in einer Epoche kunstvoll verschachtelte Sätze üblich sein, in einer anderen jedoch lange Wortbildungskonstruktionen (vgl. Schröder, 1986). Auch andere sprachspezifische Eigenschaften können die Ausprägungen der funktionalen Äquivalente zum Kodierungsbedürfnis beeinflussen; zum Beispiel gibt es im Hebräischen nur Komposita mit Substantiven oder Adjektiven als Konstituenten. Verben müssen mit anderen Mitteln kodiert werden (vgl. Clark & Berman, 1984:543). Die Auswirkungen des Kodierungsbedürfnisses beeinflussen wiederum die in den vorhergehenden Abschnitten beschriebenen Größen (vgl. 3.1 u. 3.2).

Daß die Polylexie im proportionalen Verhältnis zur Phrasenlänge und Kompositionsaktivität steht, ist bereits dargestellt worden. Wie lang jedoch eine Phrase überhaupt werden kann oder wieviele verschiedene Komposita mit einem Lexem gebildet werden können, sind Fragen, die von sprachlichen Eigenschaften entschieden werden. Dieser Zusammenhang wird in Abbildung 6 verdeutlicht.

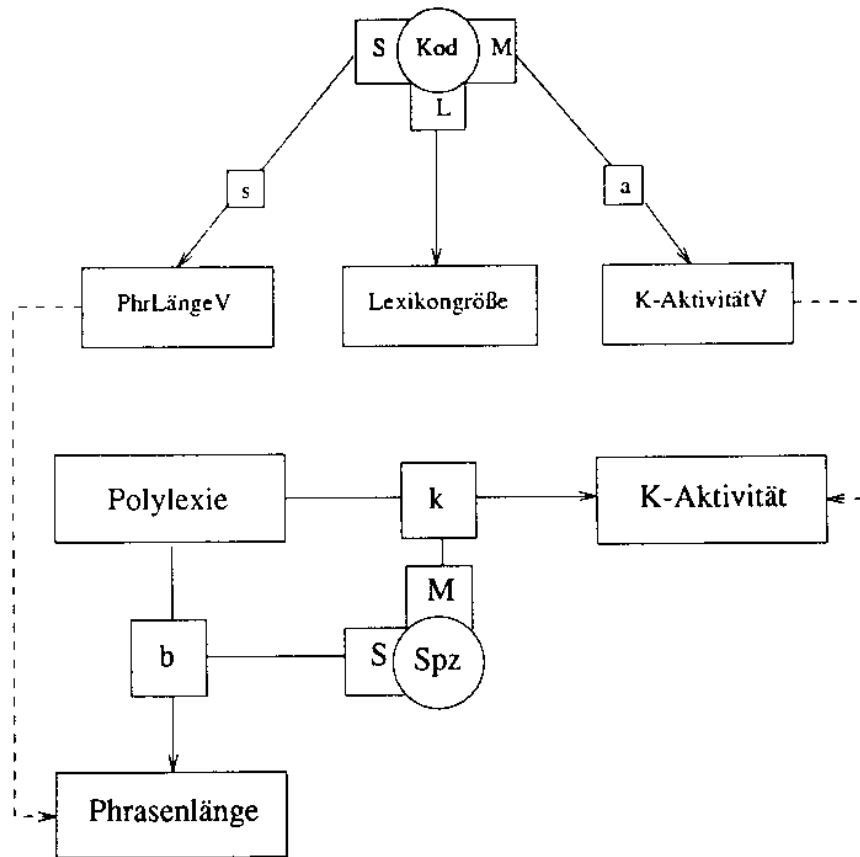


Abb. 6: Zusammenhang zwischen sprachspezifischen Eigenschaften und den Auswirkungen der Polylexie.

5. Übertragung der linguistischen Hypothesen in den statistischen Bereich

5.1 Die statistischen Hypothesen

Da der Zusammenhang zwischen Kompositionsaktivität und Polylexie analog zum Zusammenhang zwischen Polylexie und Phrasenlänge ist (vgl. 4.1), kann er entsprechend formalisiert werden:

$$(3) \quad Ka = mPl^k \quad (Ka: \text{Kompositionsaktivität}, Pl: \text{Polylexie})$$

Dabei ist der Parameter k ein Maß für den Anteil der Komposition an der Synthetizität einer Sprache. m bezeichnet die mittlere Kompositionsaktivität von Lexemen, deren Polylexie 1 beträgt. Der Zusammenhang zwischen der Polylexie und der Anzahl der Komposita im Text (TextLK-Aktivität) ist indirekt. Er kann aus der Gleichung (3) und dem folgenden Zusammenhang abgeleitet werden (vgl. 4.2):

$$(4) \quad TLKA = fKa^a \quad (TLKA: \text{Anzahl der Kompositatypen mit Lexem als Konstituente, } Ka: \text{Kompositionsaktivitat}).$$

Wird die rechte Seite der Gleichung (3) fur Ka in (4) eingesetzt, so ist das Ergebnis eine Gleichung der Form

$$(5a) \quad TLKA = fm^a Pl^{ak}$$

Das heit, die TextLK-Aktivitat ist eine Funktion der Polylexie. Wird fm^a durch g und ak durch h ersetzt, so ergibt sich die Gleichung

$$(5b) \quad TLKA = gPl^h$$

5.2 Die Operationalisierung der linguistischen Eigenschaften

5.2.1 Die Polylexie

In Anlehnung an die Kohlische Definition (vgl. 3.1) wird als Polylexie eines Lexems die Anzahl der Eintragungen unter seinem Stichwort im WDW definiert. Das Lexikon wird als die Eintrage des WDW operationalisiert. Ebenso wie bei Hammerl (1991:41) werden dabei untergeordnete Bedeutungszahlen wie 1.1, 1.1.1 nicht berucksichtigt. Als Polylexiewert eines Stichworts gilt also die grote Ordnungszahl vor einem Punkt. Die Polylexiewerte von Homonymen werden zusammengezahlt. Das jeweilige Ergebnis gilt als die Polylexie der lexikalischen Einheit.

5.2.2 Das Kompositum

Bezogen auf die obige Definition (vgl. 3.2.1) ist ein Lexem genau dann ein Kompositum, wenn es aus mehr als einer Komponente besteht, also eine Wortbildungskonstruktion ist, und wenn alle diese Komponenten als Lexeme im Lexikon verzeichnet sind. Wortbildungskonstruktionen mit Bindestrich oder Fugenelement werden von dieser Definition ebenso erfat wie Wortbildungskonstruktionen, in denen ein Verbstamm Bestandteil ist. Bindestriche und Fugenelemente werden nicht berucksichtigt; bei Verbstammen wird gepruft, ob die Infinitivform im Lexikon enthalten ist. Ausnahmen bestehen bei Morphemen wie *-schaft*, *-haft* und *-ion*; hier wird individuell entschieden, ob ein Derivat oder ein Kompositum vorliegt.

Mit Hilfe der obigen Definition wird auerdem die Klassifizierung von weiteren Wortbildungskonstruktionen bestimmt: Dies betrifft die *Zusammenbildung*,

die *Kontamination*, Bildungen mit *unikalen* und *entlehnten* Morphemen und die *Zusammenrückung*.

Zusammenbildungen wie *Gesetzgebung* und *Wichtigtuier* bezeichnet Fleischer (1983:63) als eine Form der Derivation. Bußmann (1990:870), der Duden (1984:406) und Lewandowski (1990:1282) gehen bei solchen Wortbildungskonstruktionen von einer fließenden Grenze zwischen Derivaten und Komposita aus. Nach der obigen Definition handelt es sich dann um Komposita, wenn alle Bestandteile im Lexikon enthalten sind und im anderen Fall um Derivate.

Kontaminationen wie *Euroshima* und *Sparschweinerei* sind seltene sprachliche Erscheinungen (vgl. Ortner & Ortner, 1984:101ff.; Bußmann, 1990:416; Duden, 1984:394). Sie sind - diachronisch betrachtet - eine Verschmelzung zweier Wörter. Als Komposita werden nach der Definition nur solche Bildungen angesehen, die (synchronisch) in solche Bestandteile zerlegbar sind, die im Lexikon Eintragungen besitzen.

Wortbildungskonstruktionen mit *unikalen* bzw. „blockierten“ *Morphemen* wie zum Beispiel das Wort *Himbeere* werden nach der obigen Definition nicht als Komposita klassifiziert, weil unikale Morpheme gebundene Morpheme sind.

Wortbildungskonstruktionen mit *entlehnten Morphemen* wie *Ökologie*, *Biologie*, *Fotographie* (vgl. Warren, 1990) werden nur dann als Komposita klassifiziert, wenn alle Bestandteile als freie Morpheme im Lexikon enthalten sind.

Eine weitere Form der Wortbildungskonstruktionen ist die der *Zusammenrückung*. Zusammenrückungen, die Fleischer (1983:62) als „substantivierte Sätze und Wortgruppen“ definiert, wie zum Beispiel *Gernegroß*, *Rührmichnichtan*, und die er auch auf Adverbien, wie *daher* und *dahin*, bezieht (1983:279), werden im allgemeinen als eine Sonderform der Komposition betrachtet (vgl. Bußmann, 1990:870; Lewandowski, 1990:1283; Ludwig, 1991:115). Im Duden wird dieser Begriff lediglich als zusätzliche Bezeichnung bei der Bildung von Adverbien genannt (Duden, 1984:500). Wenn zu allen Komponenten der Zusammenrückungen Einträge im Lexikon existieren, gelten auch sie nach der obigen Definition als Komposita.

Auch bei Wortbildungskonstruktionen mit Kurzwörtern oder Abkürzungen als Komponenten ist für die Klassifizierung ausschlaggebend, ob sich für diese Komponenten Lexeme im Lexikon befinden.

5.2.3 Die Kompositionsaktivität

Analog zur Bestimmung der Wortbildungsaktivität von Barz (1988:112; vgl. auch Abschnitt 2) wird die *Kompositionsaktivität* eines Lexems definiert als (Definition:) die Anzahl der Komposita, in denen das Lexem mindestens eine Konstituente ist. In dieser Untersuchung wird ein Teil der untersuchten Komposita dem Lexikon entnommen, in diesem Fall dem WDW.

5.2.4 Die TextLK-Aktivität

Die *TextLK-Aktivität* eines Lexems wird definiert als die Anzahl der Types von Komposita im Text, in denen das Lexem mindestens eine Konstituente ist. Mehrere Tokens des gleichen Kompositums werden also nicht mehrfach zur TextLK-Aktivität gezählt.

5.3 Beschreibung des Testmaterials

5.3.1 Das Wörterbuch

Das maschinenlesbare Wörterbuch, das bei dieser Untersuchung zur Bestimmung der Polylexie und der Kompositionsaktivität verwendet wird, wurde auf der Basis des WDW im Verlauf zweier Projekte an der Universität Bochum erstellt. Die Struktur der 93614 Einträge ist aus dem nachstehenden Auszug ersichtlich:

Aal	<m. 1> 1 2 3
'aa.len 1	<v.i.> 2 <refl.>
'aal.glatt	<Adj.>
'Aal.ki.ste	<f. 19>
'Aal.korb	<m. 1 u.>
'Aal.mut.ter	<f.> <f. 8u>
'Aal.quap.pe	<f. 19>
'Aal.rau.pe	<f. 19>
'Aal.tier.chen	<n. 14>
Aar	<m. 1; poet.>
Aas	1 <n. 11> 2 <n. 11; Pl. Äser> 1. 2.1. 2.2.
'Aas.blü.me	<f. 19>
'aa.sen	<v.i.>

In dem Lemma an der ersten Stelle eines Eintrags ist das Betonungsmuster und die Silbentrennung angegeben. In den spitzen Klammern sind unter anderem Angaben zur Flexion und zur Wortart enthalten. Die Zahlen außerhalb dieser Klammern kennzeichnen die unterschiedlichen Bedeutungen. Das Format läßt also eine einfache Trennung der Information zur Wortart eines bestimmten Wortes und der Information zur jeweiligen Polylexie zu. Allerdings sind die Einträge nicht immer so vollständig wie die obigen. Hauptsächlich bei Fremdwörtern fehlen bisweilen jegliche Angaben zur Wortart. Da bei dieser Untersuchung nur durchgehende Zeichenketten als Lexem betrachtet werden, werden Mehrwortlexeme wie zum Beispiel a cap'pel.la <Mus.;ital.>. ausgeschlossen.

5.3.2 Die Texte

Auf den Zusammenhang zwischen Polylexie und TextLK-Aktivität hin wird die Novelle „Ein fliehendes Pferd“ von Martin Walser (1980) untersucht. Von 3122 Types sind für die weitere Untersuchung 2807 verwertbar, d.h. sie oder alle ihre Konstituenten sind im WDW enthalten. Ferner werden 19 Texte der vdi-Nachrichten untersucht. Diese Texte entstammen einer maschinenlesbaren Ausgabe der vdi-Nachrichten⁴. Von den 3620 Types können für die weitere Untersuchung 2971 verwendet werden.

Die Texte werden nicht getrennt, sondern gemeinsam analysiert. Bei einer solchen Datenmenge, die aus verschiedenen Texten besteht, ist „Das Problem der Datenhomogenität“ (Altmann, 1992) zu berücksichtigen. Die Texte gehören zwar alle zum gleichen Genre - es handelt sich um Artikel einer Fachzeitschrift - jedoch weist jeder Text individuelle Eigenschaften auf. Für die Untersuchung ist deshalb davon auszugehen, daß der Zusammenhang zwischen Polylexie und TextLK-Aktivität für jeden Text anders ist. Das Anwendungsbedürfnis für bestimmte Lexeme ist in jedem Text sehr wahrscheinlich unterschiedlich. Deshalb ist es möglich, daß in einem Text ein bestimmtes Lexem eine sehr hohe TextLK-Aktivität besitzt, in einem anderen jedoch Bestandteil von nur wenigen Komposita ist. Auch kann der Anteil der Komposita an den Types des Textes generell unterschiedlich sein. Dies ist zum Beispiel dann der Fall, wenn das Spezifikationsbedürfnis überwiegend mit syntaktischen Mitteln bedient wird. Dennoch ist ein Zusammenhang zwischen Polylexie und TextLK-Aktivität zu erwarten, der allerdings weniger stark ausgeprägt als bei den einzelnen Texten ist. Im übrigen gilt dieses Problem auch für längere Texte, also auch für die oben beschriebene Novelle. Hier besitzen die unterschiedlichen Textabschnitte unterschiedliche Eigenschaften (vgl. Altmann, 1992:288).

5.3.3 Adhoc-Komposita aus dem SPIEGEL

Das hier untersuchte Material entstand im Rahmen einer Examensarbeit (Raether, 1987). Raether ermittelte aus vier Exemplaren der Zeitschrift *DER SPIEGEL* über 8000 Wortbildungskonstruktionen; für die vorliegende Untersuchung standen 8264 Komposita in einer Datei zur Verfügung (vgl. Raether & Rothe, 1991: 85). Die Eintragungen sind nicht im Duden-Wörterbuch aufgeführt und wurden mit dem WDW abgeglichen. Es bleiben noch 7988 Komposita, die weder im Duden-Wörterbuch, noch im WDW enthalten sind.

⁴ Es handelt sich hierbei um Texte der Ausgaben vom 12.10.90, 7.12.90, 14.12.90, 12.4.91, 10.5.91 und vom 17.5.91.

Die Komposita umfassen 6081 verschiedene Konstituenten, von denen 1987 aufgrund der Ermittlung der Einträge, die nicht im WDW enthalten sind, für die weitere Untersuchung jedoch nicht verwertbar sind. Die Textsorte, für die Wortbildungen mit Eigennamen und Abkürzungen als Konstituenten charakteristisch sind, bildet die Hauptursache für das Fehlen der Einträge im WDW. Außerdem fehlen dort auch Ziffern sowie Derivationen mit dem sogenannten „weiblich markierten Suffix 'in'“ (vgl. Frank, 1992:128ff.) wie etwa *Dolmetscherin*, *Managerin*, *Expertin*, die in den Spiegeltexten relativ häufig vertreten sind. Da das Datenmaterial lediglich die Wortbildungskonstruktionen, nicht aber sämtliche Lexeme der erfaßten Texte im *SPIEGEL* umfaßt, können hier nur Aussagen über die Komposita dieser Texte gemacht werden.⁵ Zusammenbildungen und Ableitungen wie *Beweismittelfindung* und *Schlappschwänzigkeit* sind für diese Untersuchung nicht relevant und wurden ebenso getilgt wie einige Einträge mit Orthografiefehlern. Von den verbleibenden 4096 Lexemen sind 69 Lexeme keine Konstituenten von Komposita, sondern Derivate oder Konversionen. Letztendlich können also nur 4025 Lexeme analysiert werden.

6. Empirische Überprüfung der Hypothesen

6.1 Die Schätzmethode und die Testverfahren

In den folgenden Abschnitten wird überprüft, ob die Daten mit den Hypothesen verträglich sind. Dazu werden die Hypothesengleichungen, die jeweils die Form $y = ax^b$ besitzen, logarithmiert. Es entstehen Gleichungen der Form $\ln y = \ln a + b \ln x$ (vgl. Köhler, 1986:61; Grotjahn, 1992:140).

Da diese Gleichungen lineare Funktionen darstellen, können die Parameter a und b mittels der Methode der kleinsten Quadrate geschätzt werden. Hierbei werden für die abhängigen Variablen die Mittelwerte der jeweiligen Gruppe der unabhängigen Variable verwendet. Das bedeutet allerdings, daß diese Mittelwerte um so zuverlässiger sind, je größer die ihnen zugrunde liegenden Gruppen sind und um so weniger zuverlässig, je kleiner diese Gruppen sind (vgl. Grotjahn, 1992:149f.). Damit ist eine wichtige Voraussetzung für die im folgenden verwendeten Testverfahren - nämlich die Homoskedastizität - verletzt (vgl. auch Grotjahn, 1992:126, 150). Die Schätzmethode der kleinsten Quadrate wird von Grotjahn (1992:150) für einen solchen Fall als nicht effizient bezeichnet.

Außerdem ist zu beachten, daß die Schätzung sich auf die linearisierten und nicht auf die nichtlinearen Gleichungen bezieht. Das bedeutet auch, daß die Testverfahren nur etwas über die Anpassung der Daten an die linearisierten Gleichungen aussagen (vgl. Grotjahn, 1992:140, 156). Jedoch ist nach Grotjahn

⁵ Das Problem der vorsortierten Stichprobe wurde bereits in der Einleitung aufgezeigt.

(1992:142) die Linearisierung trotz aller Kritikpunkte in vielen Fällen „a very convenient method which yields a good approximation to the original nonlinear model as fitted with the help of more sophisticated methods.“ In diesem Sinne werden zwei Testverfahren verwendet, um das Ergebnis dieser Schätzung bewerten zu können - wenn auch mit den dargestellten Einschränkungen:

Das erste Testverfahren ist der *F*-Test, der ein Maß dafür ist, ob die Abweichungen der theoretischen Werte von den tatsächlichen Werten so gering sind, daß die Schätzung als richtig angenommen werden kann (vgl. Köhler, 1986:99; Tränkle, 1980:185ff.).

Das zweite Verfahren ist die Berechnung des Determinationskoeffizienten (vgl. Grotjahn, 1992:152ff.; Tränkle, 1980:84f.). Der Determinationskoeffizient gibt den Anteil der durch die Prädiktorvariable erklärte Varianz der theoretischen Daten an der Gesamtvarianz an. Grotjahn (1992:153) empfiehlt die Verwendung des Ausdrucks

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

für den Determinationskoeffizienten. Dabei bezeichnet y_i jeweils den empirischen und \hat{y}_i den theoretischen Wert. \bar{y} bezeichnet den arithmetischen Durchschnitt. Ferner wird auch der adjustierte Determinationskoeffizient

$$AR^2 = 1 - a(1 - R^2) \quad \text{mit} \quad a = (n-1)/(n-J-1)$$

(wobei n : Umfang der Stichprobe, J : Anzahl der Prädiktorvariablen) angegeben (vgl. Grotjahn, 1992:155f.). Die Berechnungen wurden mit SPSS for Windows 5.01 durchgeführt.

6.2 Polylexie und Kompositionsaktivität

Im folgenden wird die Hypothese $Ka = mPl^k$ (vgl. 5.1) überprüft. Für die Anpassungen der Funktion an die Mittelwerte wird die folgende Einschränkung gemacht: Sobald Klassengrößen mit zwei oder weniger Vorkommen auftreten, werden die ihnen zugehörigen und die ihnen folgenden Mittelwerte nicht mehr für die Anpassungen berücksichtigt. Der für die Hypothese interessierende Zusammenhang zwischen Polylexie und Kompositionsaktivität wird aus der durchschnittlichen Kompositionsaktivität der Polylexieklassen ersichtlich (vgl. Tabelle 1). Mit steigender Polylexieklasse steigt auch die Kompositionsaktivität. Allerdings ist ab der sechsten Polylexieklasse ein Einschnitt in dieser Steigung zu erkennen. Die

Aufteilung nach den Wortarten Nomina, Verben und Adjektive liefert ein ähnliches Bild (vgl. Tabellen 2 bis 4).

Tabelle 1
Anpassung der Funktion $Ka = mPl^k$ an den WDW

<i>Polylexie</i>	<i>Mittlere K-Aktivität</i>		<i>Absolute Häufigkeit</i>
	<i>beobachtet</i>	<i>berechnet</i>	
1	0.3	0.82	83493
2	3.23	2.99	7068
3	10.14	6.38	1769
4	17.51	10.92	634
5	30.41	16.56	283
6	27.94	23.27	150
7	50.13	31.03	92
8	68.75	39.82	40
9	46.58	49.61	19
10	36.46	60.40	13
11	52.64	72.16	14
12	71.50	84.88	6
13	43.83	98.57	6
$m = 0.821$ $k = 1.867$	$F(1,9) = 82.003$ $R^2 = 0.883$		$SigF < 0.000049$ $AR^2 = 0.871$

Tabelle 2
Anpassung der Funktion $Ka = mPl^k$ an den WDW (Nomen)

<i>Polylexie</i>	<i>Mittlere K-Aktivität</i>		<i>Absolute Häufigkeit</i>
	<i>beobachtet</i>	<i>berechnet</i>	
1	0.37	0.66	58288
2	4.91	3.58	3555
3	14.95	9.66	831
4	25.38	19.53	269
5	50.45	33.72	128
6	44.51	52.68	61
7	69.58	76.81	31
8	71.60	106.49	10
9	119.00	142.06	5
$m = 0.657$	$F(1,7) = 160.908$		$SigF < 0.000049$

$k = 2.447$	$R^2 = 0.958$	$AR^2 = 0.952$
-------------	---------------	----------------

Tabelle 3
Anpassung der Funktion $Ka = mPl^k$ an den WDW (Verben)

Polylexie	Mittlere K-Aktivität		Absolute Häufigkeit
	beobachtet	berechnet	
1	0.25	0.33	8138
2	1.24	1.18	1982
3	2.52	2.49	579
4	4.53	4.23	238
5	6.78	6.39	78
6	15.83	8.95	54
7	11.72	11.89	43
8	17.89	15.21	19
9	18.90	18.91	10
10	19.63	22.97	8
11	22.86	27.38	7
12	33.33	32.15	3
13	26.60	37.27	5
$m = 0.328$ $k = 1.84$	$F(1,11) = 432.865$ $R^2 = 0.975$		$SigF < 0.000049$ $AR^2 = 0.973$

Tabelle 4
Anpassung der Funktion $Ka = mPl^k$ an den WDW (Adjektive)

Polylexie	Mittlere K-Aktivität		Absolute Häufigkeit
	beobachtet	berechnet	
1	0.19	0.36	10882
2	1.60	1.88	926
3	9.45	4.96	239
4	28.90	9.87	80
5	17.85	16.85	47
6	25.75	26.07	20
7	53.83	37.71	12
8	24.67	51.93	6
9	44.50	68.85	2
10	76.33	88.61	3

$m = 0.357$	$F(1,8) = 76.787$	$SigF < 0.000049$
$k = 2.395$	$R^2 = 0.906$	$AR^2 = 0.894$

Die Abbildungen 7 und 8 stellen die Mittelwerte der Einträge des gesamten WDW und der einzelnen Wortarten dar.

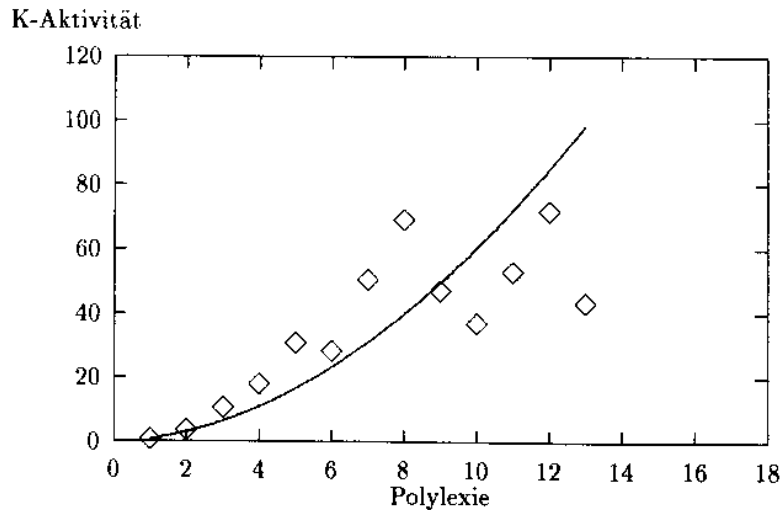


Abb. 7: Die Kompositionsaktivität in Abhängigkeit von der Polylexie (alle Wortklassen)

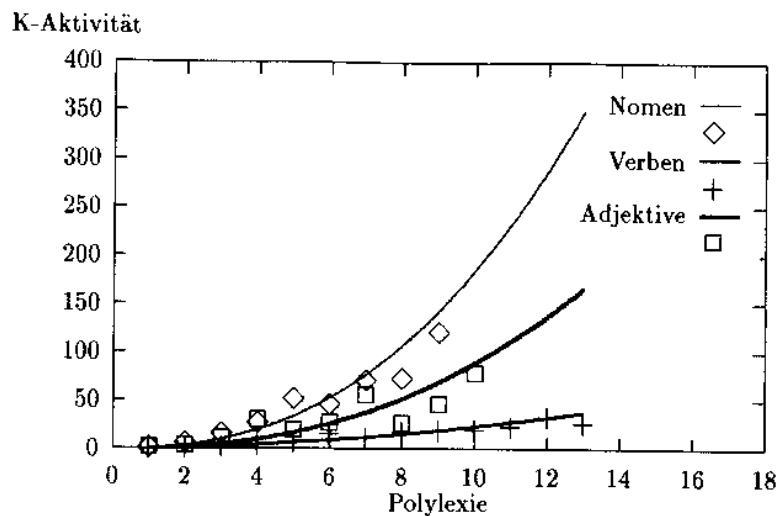


Abb. 8: Die Kompositionsaktivität in Abhängigkeit von der Polylexie (Nomina, Verben, Adjektive)

Die Schwankungen bei den größeren Polylexiewerten lassen sich anhand der relativ kleinen Klassengrößen und dem Auftreten weniger stark nach oben abweichender Werte erklären.

Wie aus den Tabellen 1 bis 4 ersichtlich ist, sind sowohl die Werte für die Determinationskoeffizienten als auch die Signifikanzniveaus für die F -Tests hoch. Deshalb kann die obige Hypothese für den gesamten WDW und für die einzelnen

Wortartklassen als bestätigt angesehen werden. Es ist nicht nur interessant, daß die durchschnittlichen Kompositionsaktivitäten der jeweiligen Polylexieklassen sich tatsächlich so verhalten wie es die Hypothese vorhersagt. - Interessant ist vielmehr auch, wie der Anteil der Wörter mit einer bestimmten Polylexie für die einzelnen Kompositionsaktivitäten verteilt ist. Daß beispielsweise die absoluten Häufigkeiten für die Kompositionsaktivität 2 bei der Polylexieklasse 11151 Lexeme und bei der Polylexieklasse 2 nur 371 Lexeme beträgt, bedeutet keineswegs, daß hier mit steigender Polylexie die Kompositionsaktivität sinkt. Vielmehr spielt hier die Tatsache eine Rolle, daß mit der steigenden Polylexie die Klassengröße sinkt. Die erste Polylexieklasse umfaßt 83493 Lexeme, die zweite 7068 Lexeme. Um den Zusammenhang zwischen den einzelnen Werten der Polylexie- und Kompositionsaktivitätsklassen vergleichen zu können, müssen die relativen Häufigkeiten bezüglich der Klassengröße der Polylexie berechnet werden.

Es zeigen sich zwei unterschiedliche Tendenzen. Die erste betrifft die Klassen mit einer niedrigen Kompositionsaktivität: Hier sinkt der relative Anteil der Polylexie mit dem Anstieg des Polylexiewertes. Es gibt also relativ wenige Lexeme mit einer niedrigen Kompositionsaktivität und einer hohen Polylexie. Der umgekehrte Zusammenhang zeigt sich bei hohen Kompositionsaktivitäten: Relativ viele Lexeme haben eine hohe Polylexie. Der Anteil steigt, je größer die Polylexie ist.

Die Abbildungen 9 bis 11 veranschaulichen diese beiden Tendenzen. Die ersten zwei Kurven sind konvex fallend, die nächsten vier konkav steigend und alle weiteren in immer stärkerem Maße konvex steigend (vgl. auch Altmann & Schwibbe, 1989:6). Der inverse Zusammenhang zwischen Polylexie und Kompositionsaktivität in den ersten Klassen der Kompositionsaktivität kehrt sich immer weiter um in den positiven Zusammenhang der oberen Klassen der K-Aktivität. Die Mittelwerte, an die die Hypothesengleichung bereits angepaßt wurde, geben also nur ein relativ grobes Bild des Zusammenhangs zwischen Polylexie und Kompositionsaktivität wieder.

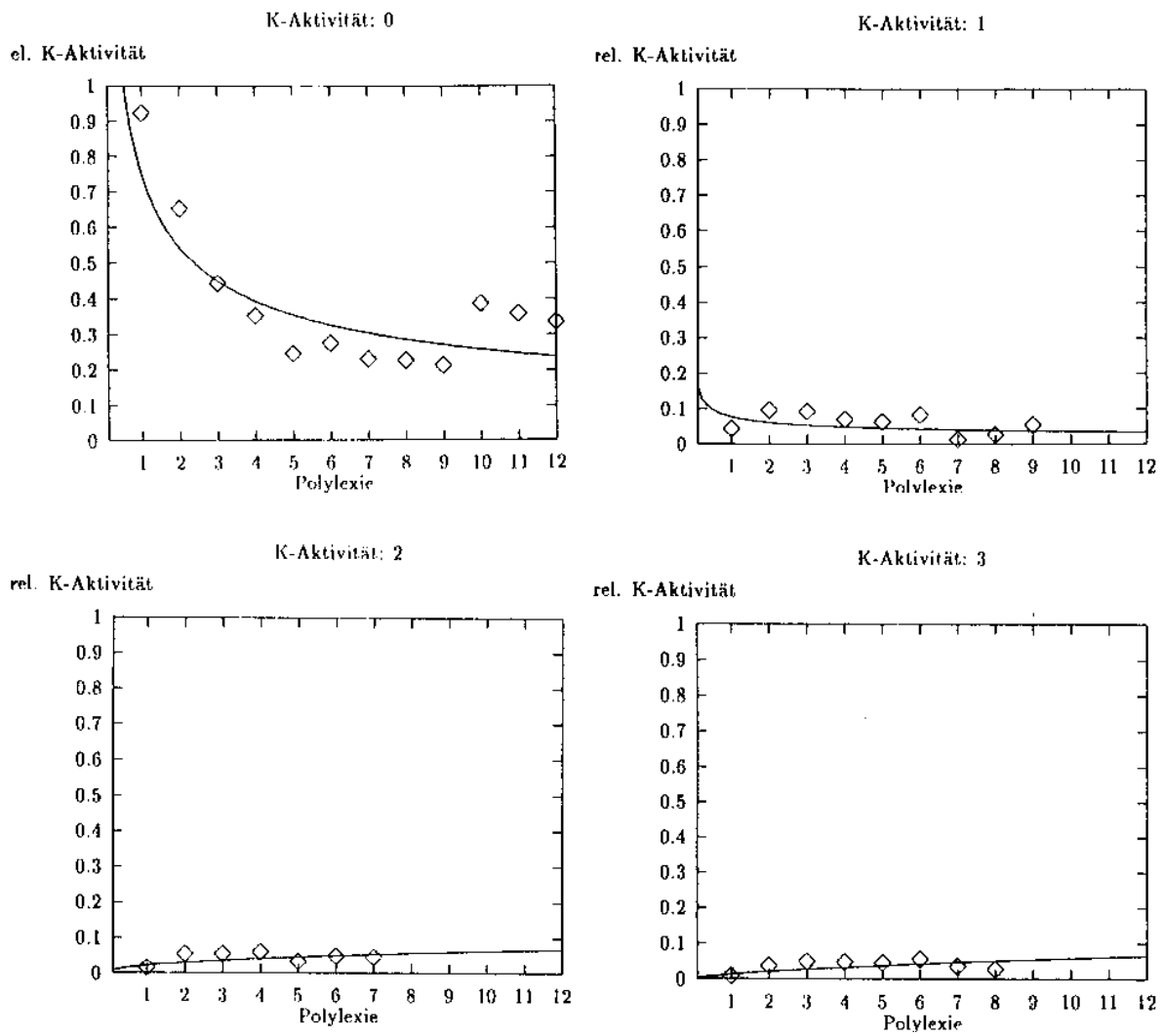


Abb. 9: Die relativen Häufigkeiten der Polylexiewerte in den ersten vier Klassen der *K*-Aktivität

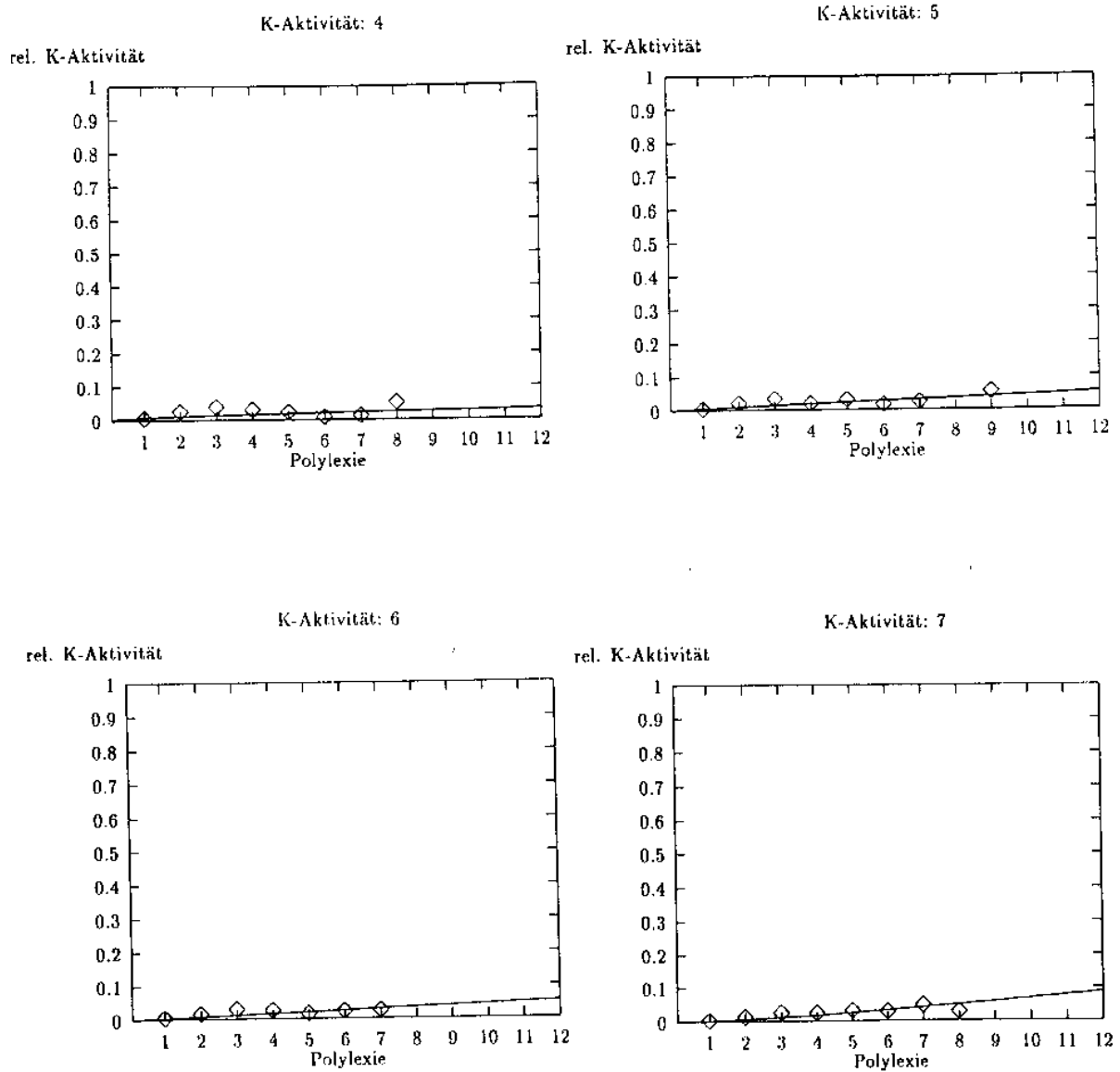


Abb. 10: Die relativen Häufigkeiten der Polylexiewerte in der fünften bis achten Klasse der K-Aktivität

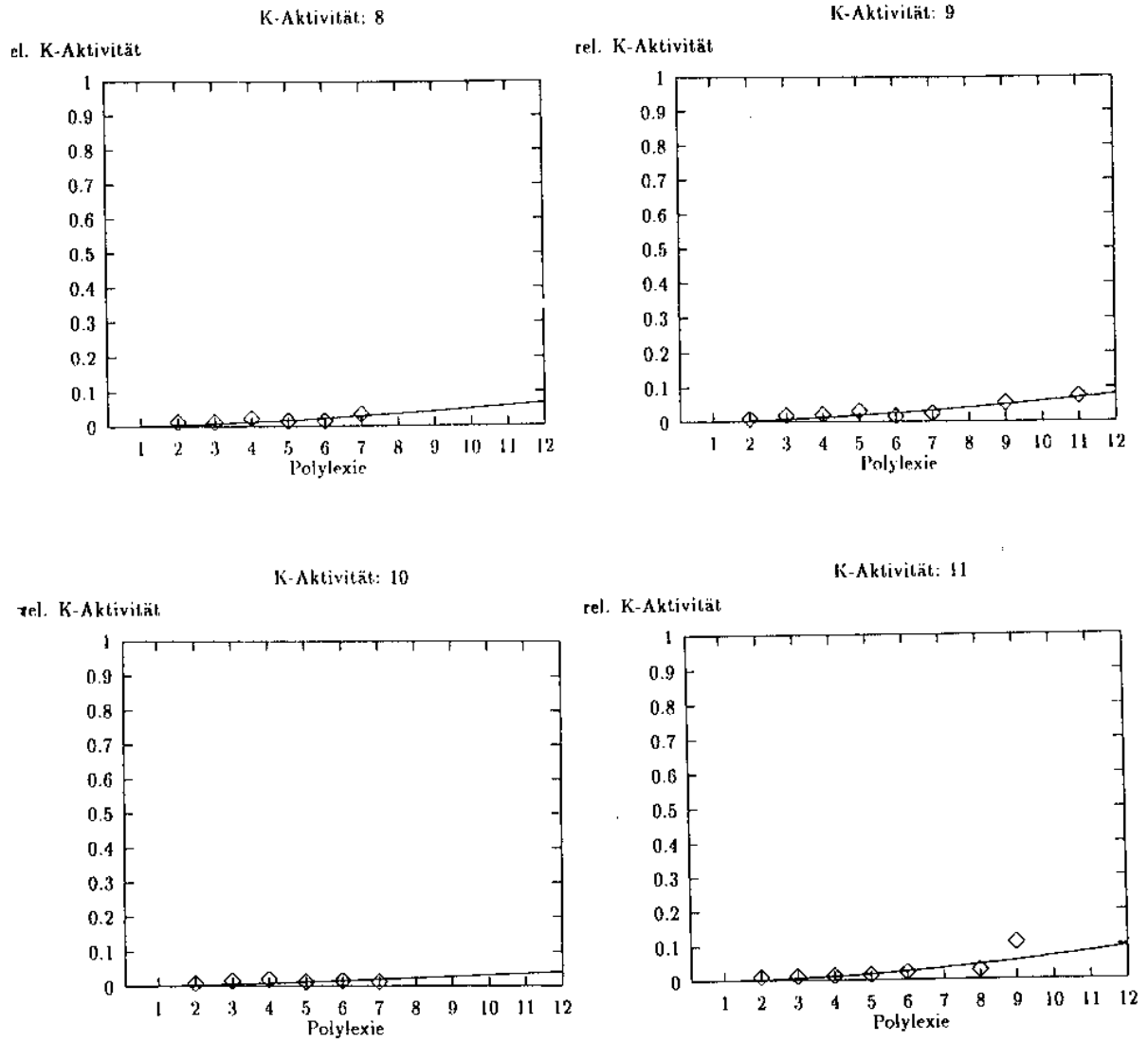


Abb. 11: Die relativen Häufigkeiten der Polylexiewerte in der neunten bis zwölften Klasse der K-Aktivität.

Die Tabelle 5 stellt die Anpassungen der Exponentialfunktion für jede Zeile der der Tabelle mit den relativen Häufigkeiten dar.

Tabelle 5

Anpassungen der Funktionen $RKa_i = aPl^b; i = 0..11$ an die ersten 12 K-Aktivitätsklassen des WDW

<i>i</i>	<i>a</i>	<i>b</i>	<i>FG</i>	<i>F</i>	<i>SigF</i>	<i>R</i> ²	<i>AR</i> ²
0	0.7480	-0.461	1,10	15.188	0.0030	0.603	0.563
1	0.0783	-0.331	1,7	0.914	0.3709	0.116	0.108
2	0.0228	0.443	1,5	2.911	0.1487	0.368	0.242
3	0.0137	0.618	1,6	4.255	0.0847	0.415	0.317
4	0.0086	0.520	1,6	1.236	0.3088	0.171	0.033
5	0.0056	0.894	1,6	8.005	0.0300	0.572	0.500
6	0.0039	1.040	1,5	8.767	0.0315	0.637	0.564
7	0.0025	1.411	1,6	25.576	0.0023	0.810	0.778
8	0.0017	1.466	1,5	18.772	0.0075	0.790	0.748
9	0.0014	1.601	1,7	48.652	0.0002	0.874	0.856
10	0.0015	1.314	1,5	11,335	0.0200	0.693	0.633
11	0.0009	1.867	1,6	43.437	0.0006	0.879	0.858

6.3 Polylexie und TextLK-Aktivität

Die Hypothese $TLKA = gPl^h$ (vgl. 5.1) wird im folgenden an drei Stichproben überprüft⁶. Das Verfahren zur Gewinnung der Datenmenge ist analog zu dem bei der vorherigen Anpassung. Zunächst soll die Funktion an die mittleren TextLK-Aktivitäten der Polylexieklassen bei der Novelle, sowie bei den Zeitungstexten angepaßt werden. Die Tabellen 6 und 7 sowie die Abbildungen 12 und 13 stellen die Ergebnisse der Anpassungen an die Daten dar.

Die Signifikanzniveaus für die *F*-Tests sind sehr hoch; aufgrund dessen können die Hypothesen angenommen werden. Die Werte für den Parameter *h* sind um einiges niedriger als die Werte für den Exponenten *k* bei der Anpassung der Funktion $Ka = mPl^k$ an die Daten aus dem WDW (vgl. 6.2). Dies läßt sich dadurch erklären, daß die Polylexie und die TextLK-Aktivität nur indirekt zusammenhängen und der Parameter *h* das Produkt aus dem *k* der vorherigen Hypothese und einem Parameter *a* ist, der vom Anwendungsbedürfnis mitbestimmt wird (vgl. 5.1). Wenn für *k* der Wert 1.867 angenommen wird (vgl. 6.2, Tabelle 1) und *h* den Wert 1.005 besitzt (vgl. Tabelle 6), dann besitzt der Parameter *a* den Wert 0.538. Wird zusätzlich angenommen, daß der Parameter *m* in

⁶ vgl. 5.3.2 und 5.3.3.

$Ka = mPl^k$ den Wert 0.821 besitzt, so läßt sich f in $TLKA = fKa^a$ (vgl. (4)) berechnen. Für $g = 0.290$ besitzt dieser Parameter den Wert 0.322. Dieser Wert besagt, daß Lexeme mit einer Kompositionsaktivität von 1 im Durchschnitt nur eine höchstens dreimal so kleine TextLK-Aktivität besitzen. Das gilt allerdings nur für die Lexeme der analysierten Novelle, für die diese Berechnung durchgeführt wurde. Für die analysierten vdi-Nachrichten ist $a = 0.507$ und $f = 0.765$. Bei der Berechnung dieser Werte sind allerdings die Fehler bei der Berechnung von g und h mit eingeflossen. Bei einer Anpassung der Funktion $TLKA = fKa^a$ an die Daten würden sicherlich andere Werte ermittelt. Diese würden aber, was ihr Vorzeichen und ihre Größenordnung betrifft, den eben errechneten Parametern gleichen (vgl. Köhler, 1986:113).

Tabelle 6

Anpassung der Funktion $TLKA = gPl^h$ an die Daten der Novelle „Ein fliehendes Pferd“

<i>Polylexie</i>	<i>Mittlere TLK-Aktivität</i>		<i>Absolute Häufigkeit</i>
	<i>beobachtet</i>	<i>berechnet</i>	
1	0.38	0.29	1048
2	0.45	0.58	724
3	0.87	0.87	458
4	1.10	1.17	231
5	1.25	1.46	136
6	1.70	1.75	79
7	2.27	2.05	60
8	4.23	2.34	26
9	1.42	2.64	12
10	1.67	2.93	9
11	4.75	3.23	8
12	4.25	3.52	4
13	2.60	3.82	5
14	7.00	4.11	3
$g = 0.290$ $h = 1.005$	$F(1,12) = 50.865$ $R^2 = 0.809$		$SigF < 0.000049$ $AR^2 = 0.793$

Tabelle 7

Anpassung der Funktion $TLkA = gPl^h$ an die Daten der VDI-Nachrichten

Polylexie	Mittlere TLK-Aktivität		Absolute Häufigkeit
	beobachtet	berechnet	
1	0.74	0.69	1244
2	1.03	1.33	775
3	2.17	1.96	411
4	2.49	2.57	215
5	3.98	3.18	124
6	2.77	3.78	77
7	6.04	4.37	55
8	7.85	4.96	26
9	5.10	5.55	10
10	2.89	6.13	9
11	8.44	6.71	9
12	5.75	7.28	4
13	10.20	7.86	5
$g = 0.692$		$F(1,11) = 53.610$	$SigF < 0.000049$
$h = 0.947$		$R^2 = 0.829$	$AR^2 = 0.814$

TextLK-Aktivität

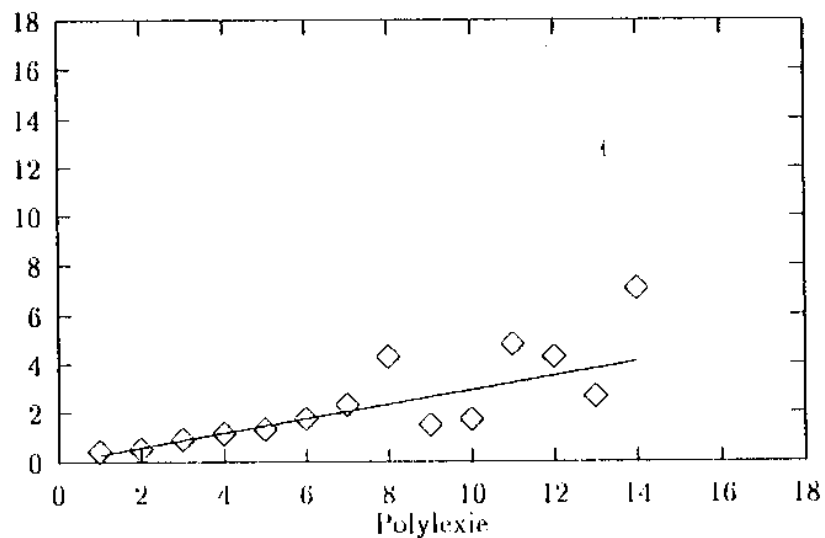


Abb. 12: Die durchschnittliche TextLK-Aktivität in Abhängigkeit von der Polylexie (Ein fliehendes Pferd)

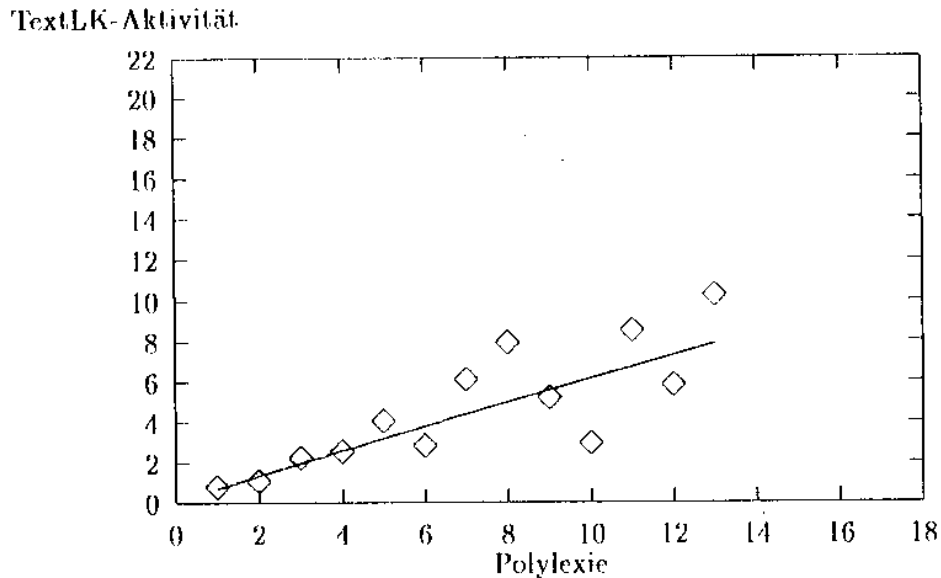


Abb. 13: Die durchschnittliche TextLK-Aktivität in Abhängigkeit von der Polylexie (VDI-Nachrichten)

Die Komposita aus den Texten des SPIEGEL müssen mit dem Vorbehalt untersucht werden, daß es sich hier um eine vorsortierte Stichprobe handelt (vgl. 5.3.3).

Die Anpassung der Funktion $TLKA = gP^h$ an die durchschnittlichen Text-LK-Aktivitäten der einzelnen Polylexieklassen kann zu einem Vergleich mit den anderen Texten nur den Anpassungen an die jeweiligen Teilmengen der *Komposita* aus diesen Texten gegenübergestellt werden. Die Tabellen 8 bis 10 sowie die Abbildungen 14 bis 16 stellen die Anpassungen der Funktion an diese Daten dar.

Aufgrund der niedrigen Signifikanzniveaus für die F -Tests kann die Hypothese auch für die Komposita aus dem SPIEGEL angenommen werden. Die Werte für den Faktor sind größer und die Werte für den Exponenten kleiner, wenn lediglich die Stichproben der Komposita untersucht werden. Bei der Stichprobe aus dem SPIEGEL ist die TextLK-Aktivität, die ein Lexem mit der Polylexie 1 und der Eigenschaft, daß es Komposita bildet, im Durchschnitt besitzt, größer als bei den anderen Stichproben. Das läßt sich einerseits durch die textspezifischen Eigenschaften der Artikel aus dem Nachrichten-Magazin erklären. Eine andere Erklärung ist, daß es sich hier um eine Stichprobe nichtusueller Komposita handelt und daß die Konstituenten solcher Komposita andere Eigenschaften hinsichtlich der Polylexie besitzen als die Konstituenten beliebiger Komposita.

Tabelle 8
Anpassung der Funktion $TLKA = gPl^h$ an Textdaten ohne Simplizia (SPIEGEL)

Polylexie	Mittlere TLK-Aktivität		Absolute Häufigkeit
	beobachtet	berechnet	
1	2.36	2.48	2057
2	3.35	3.64	998
3	4.50	4.55	454
4	5.19	5.34	227
5	7.32	6.04	119
6	8.19	6.69	62
7	8.82	7.28	49
8	7.09	7.84	23
9	9.50	8.37	10
10	5.29	8.87	7
11	10.13	9.35	8
$g = 2.477$ $h = 0.554$	$F(1,9) = 36.029$ $R^2 = 0.800$		$SigF = 0.0002$ $AR^2 = 0.778$

Tabelle 9
Anpassung der Funktion $TLKA = gPl^h$ an die Daten der Novelle „Ein fliehendes Pferd“ ohne Simplizia

Polylexie	Mittlere TLK-Aktivität		Absolute Häufigkeit
	beobachtet	berechnet	
1	1.38	1.32	291
2	1.45	1.82	224
3	2.27	2.19	175
4	2.52	2.50	101
5	2.50	2.78	68
6	3.44	3.02	39
7	4.39	3.24	31
8	6.11	3.45	18
9	2.13	3.64	8
10	2.50	3.82	6
11	6.33	4.00	6
12	4.25	4.16	4
13	3.25	4.32	4
$g = 1.318$ $h = 0.462$	$F(1,11) = 13.083$ $R^2 = 0.543$		$SigF = 0.0040$ $AR^2 = 0.502$

Tabelle 10

Anpassung der Funktion $TLKA = gPl^h$ an die Daten der VDI-Nachrichten ohne Simplizia

Polylexie	Mittlere TLK-Aktivität		Absolute Häufigkeit
	beobachtet	berechnet	
1	1.90	1.84	486
2	2.18	2.85	367
3	3.92	3.68	228
4	4.47	4.40	120
5	6.33	5.07	78
6	4.73	5.68	45
7	9.76	6.26	34
8	11.33	6.81	18
9	5.67	7.33	9
10	3.71	7.83	7
11	10.86	8.31	7
12	5.75	8.78	4
13	12.75	9.23	4
$g = 1.843$ $h = 0.628$		$F(1,11) = 19.131$ $R^2 = 0.635$	$SigF = 0.0011$ $AR^2 = 0.602$

TextLK-Aktivität

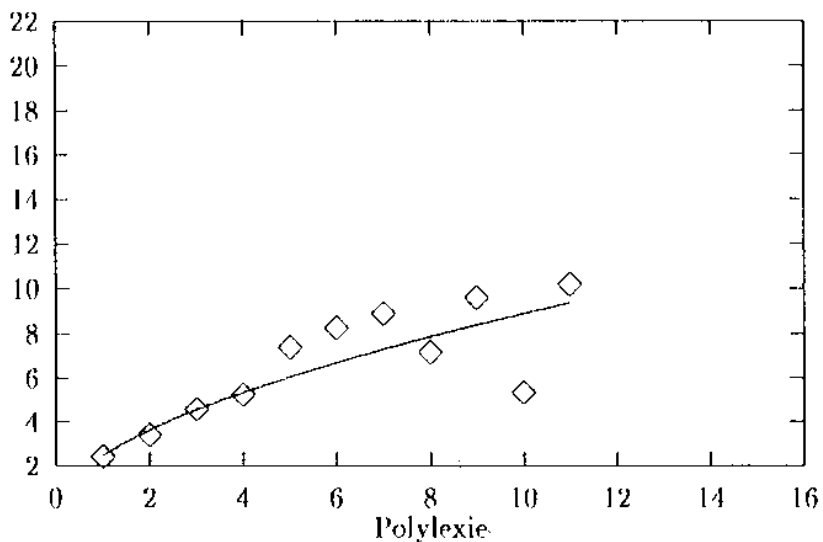


Abb. 14: Die durchschnittliche TextLK-Aktivität in Abhängigkeit von der Polylexie (SPIEGEL, Komposita)

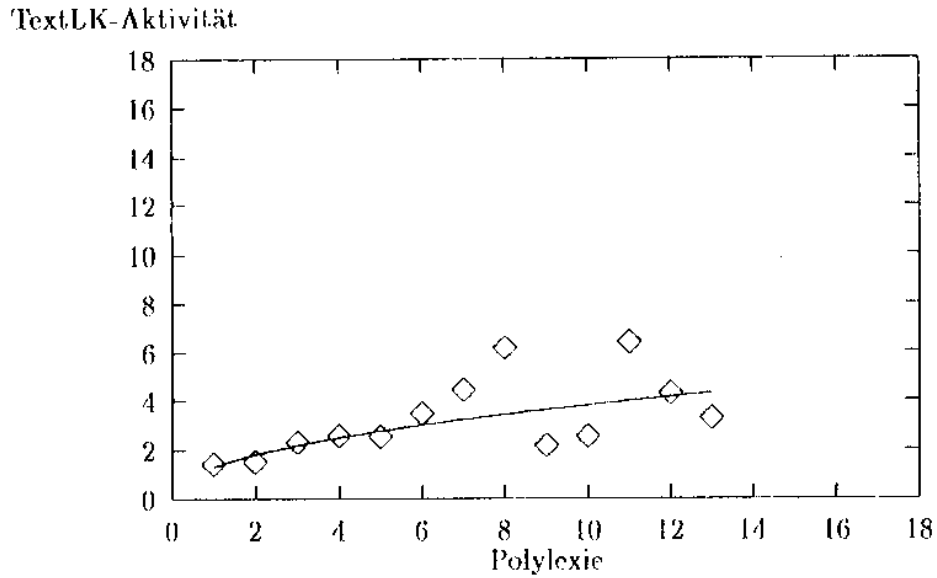


Abb. 15: Die durchschnittliche TextLK-Aktivität in Abhängigkeit von der Polylexie (Ein fliehendes Pferd, ohne Simplizia)

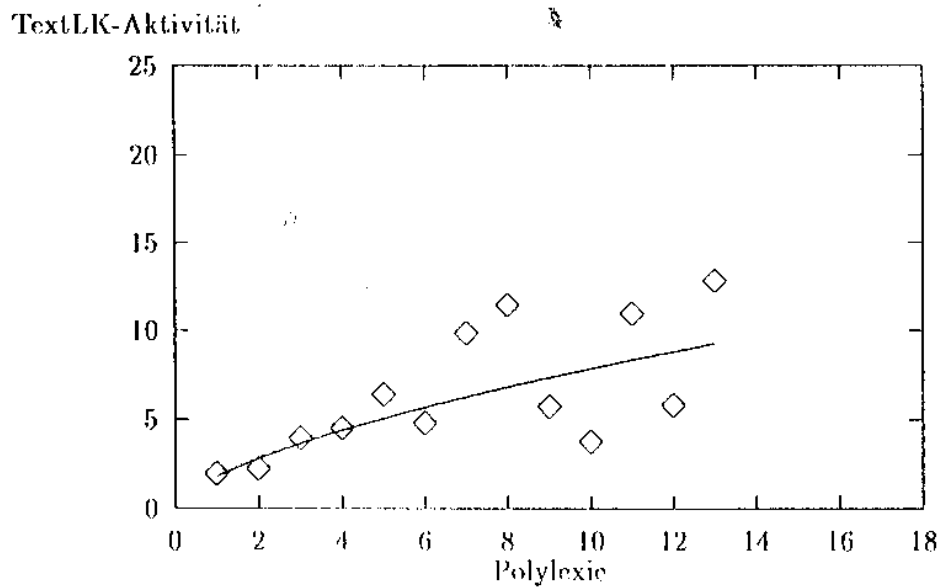


Abb. 16: Die durchschnittliche TextLK-Aktivität in Abhängigkeit von der Polylexie (VDI-Nachrichten, ohne Simplizia)

7. Zusammenfassung und Diskussion der Ergebnisse

Die durchgeführten Tests bestätigen die Verträglichkeit der Daten mit der Hypothese $Ka = mPl^k$ und der Hypothese $TLKA = gPl^h$.

Die Signifikanzniveaus liegen bei der Anpassung der ersten Funktion an die Daten unter dem Wert 0.000049. Sie sind also hochsignifikant. Auch die Signifikanzniveaus bei der Anpassung der zweiten Funktion mit Werten unter 0.000049 sind hoch. Die Anpassungen der beiden Funktionen an Daten, denen nur aus Komposita bestehende Stichproben zugrundeliegen, bestätigen die Hypothesen hinsichtlich dieser Stichproben.

Es wurde auch dargestellt, daß sich die Klassen der Kompositionsaktivität und der TextLK-Aktivität hinsichtlich der relativen Häufigkeit der Polylexie-klassen unterscheiden. In den unteren Klassen der Kompositions-aktivität und der TextLK-Aktivität ist der Anteil der Lexeme mit der Polylexie 1 am größten und sinkt mit steigender Polylexie. Je größer die Klasse der K- oder TextLK-Aktivität ist, desto mehr kehrt sich dieser Zusammenhang um. Außerdem zeigen zwei Vergleiche mit Stichproben, die nur aus Komposita bestehen, daß die Anpassungen zu diesen Stichproben wesentlich andere Werte für die Parameter liefern als die Anpassungen an die Daten auf der Basis aller Lexeme. Diese Beobachtungen und Ergebnisse bestätigen nicht nur die beiden Hypothesen, sondern weisen außerdem auch darauf hin, daß die Verteilung der Polylexie im Bereich der Lexik bestimmten Tendenzen unterliegt.

8. Ausblick

Die vorliegende Untersuchung zeigt, daß sowohl die Kompositionsaktivität in der Lexik als auch die TextLK-Aktivität in Texten von der Polylexie abhängen. Sie liefert damit eine Erklärung für das Auftreten und die Ausprägungen dieser beiden sprachlichen Eigenschaften bei den Lexemen der deutschen Sprache. Auf diese Weise leistet sie einen kleinen Beitrag zum Forschungsbereich der synergetischen und der quantitativen Linguistik. Jedoch bestehen im Bereich der quantitativen Kompositaforschung bislang mehr offene Fragen als Erklärungen. Einige dieser Fragen werden im folgenden zusammengestellt.

Hinsichtlich der Verteilung der Polylexie fällt auf, daß die Zahl der Lexeme mit steigender Polylexie abnimmt. Für diese Erscheinung muß es eine Erklärung geben, die möglicherweise im Bereich der Diversifikation zu finden ist. Damit zusammenhängend könnte die Frage nach der zugrundeliegenden Verteilung geklärt werden.

Etliche der Altmannschen „Hypothesen about Compounds“ wurden noch nicht überprüft. Die Hypothesen, die die semantische Korrespondenz zwischen den Konstituenten und den jeweiligen Simplizia sowie die semantische Korrespondenz zwischen Konstituenten und Kompositum betreffen, sind nach Altmann

(1989:103) schwierig zu messen, weil es (bei zweigliedrigen Komposita) nur drei Möglichkeiten der (Nicht-)Übereinstimmung gibt. Möglicherweise ist es sinnvoll, die Stärke dieser (Nicht-)Übereinstimmung als Maß für die Untersuchung zu verwenden, wie dies Barz (1988) tut. Dann wäre nicht zu messen, *ob* oder *ob nicht* eine Übereinstimmung zwischen der Bedeutung einer Konstituente und ihrer ursprünglichen Bedeutung besteht; sondern die Fragestellung wäre, wie stark diese Übereinstimmung ist, das heißt, wieviel gemeinsame Sememe vorhanden sind. Es ist anzunehmen, daß die Zahl der Komposita proportional mit der Zahl der gemeinsamen Sememe ihrer Konstituenten und deren Einträgen im Lexikon steigt. Ebenso ist anzunehmen, daß die Zahl der gemeinsamen Sememe von Konstituenten und Komposita proportional mit der Länge der Komposita zunimmt. Also könnten diese Hypothesen nach einer Veränderung des Maßes für die semantische Korrespondenz auch mit Stichproben kurzer Komposita getestet werden.

Interessant ist auch die Frage, inwiefern die Prototypizität die Kompositionsaktivität bestimmt. Für ihre Beantwortung müßten Daten aus dem Bereich der Prototypforschung herangezogen werden. Außerdem läßt diese Untersuchung die Frage offen, ob die Menge der nichtusuellen Komposita auf andere Weise von der Polylexie der Konstituenten abhängt als die Menge der usuellen Komposita. Schließlich sollten die dargestellten Fragestellungen und Hypothesen auch auf verschiedene Sprachen bezogen und daran überprüft werden. Solche Untersuchungen könnten allgemeinsprachliche Gesetze zum Ergebnis haben, die wiederum in eine linguistische Theorie integriert werden können.

Literatur

- Altmann, G.** (1989). Hypotheses about Compounds. In R. Hammerl (Hg.), *Glottometrika 10* (S. 100-107), Bochum: Brockmeyer.
- Altmann, G.** (1992). Das Problem der Datenhomogenität. In B. Rieger (Hg.), *Glottometrika 13* (S. 287-298), Bochum: Brockmeyer.
- Altmann, G., Beöthy, E., & Best, K.-H.** (1982). Die Bedeutungskomplexität der Wörter und das Menzerathsche Gesetz. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationswissenschaft*, 35.5, 537-543.
- Altmann, G., Best, K.-H., & Kind, B.** (1987). Eine Verallgemeinerung der semantischen Diversifikation. In I. Fickermann (Hg.), *Glottometrika 8* (S. 130-139), Bochum: Brockmeyer.
- Altmann, G., & Lehfeldt, W.** (1980). *Einführung in die quantitative Phonologie*. Bochum: Brockmeyer.
- Altmann, G., & Schwibbe, M.H.** (1989). *Das Menzerathsche Gesetz in informationsverarbeitenden Systemen*. Hildesheim, Zürich, New York: Olms.
- Augst, G.** (1975). *Lexika zur Wortbildung. Morpheminventar*. 3 Bde. Tübingen: Narr.
- Barz, I.** (1982). Motivation und Wortbildungsbedeutung. Eine Diskussion sowjetischer Forschungsergebnisse. *Beiträge zur Erforschung der deutschen Sprache*, 2, 5-21.
- Barz, I.** (1987). Zur Wortbildungsaktivität als Eigenschaft primärer Lexeme. *Beiträge zur Erforschung der deutschen Sprache*, 7, 5-11.
- Barz, I.** (1988). *Nomination durch Wortbildung*. Leipzig: VEB Verlag Enzyklopädie.
- Barz, I.** (1989). Determinanten der Wortbildungsaktivität primärer Lexeme. *Beiträge zur Erforschung der deutschen Sprache*, 9, 32-36.
- Bußmann, H.** (1990). *Lexikon der Sprachwissenschaft*. Stuttgart: Kröner.
- Clark, E.V., & Berman, R.A.** (1984). Structure and Use in the Acquisition of Word Formation. *Language*, 60.3, 542-590.
- Cortes, J. De** (1982). Nominale Komposita aus drei oder mehr Konstituenten: Bemerkungen zu der Komposition in der wissenschaftlichen Fachsprache der Wirtschaft. *Fachsprache*, 4.1, 18-31.
- Dederding, H.-M.** (1982). *Wortbildung, Syntax, Text. Nominalkomposita und entsprechende syntaktische Strukturen in deutschen Patent- und Auslegungsschriften*. Erlangen: Palm & Enke.
- Dederding, H.-M.** (1983). Wortbildung und Text (Zur Textfunktion von Nominalkomposita). *Zeitschrift für germanistische Linguistik*, 11, 49-64.
- Duden** (1984). *Grammatik der deutschen Gegenwartssprache* Mannheim, Wien, Zürich: Bibliographisches Institut/Dudenverlag.

- Fickermann, I., Markner-Jäger, B., & Rothe, U.** (1984). Wortlänge und Bedeutungskomplexität. In J. Boy & R. Köhler (Hg.), *Glottometrika 6* (S. 115-126), Bochum: Brockmeyer.
- Fleischer, W.** (1983). *Wortbildung der deutschen Gegenwartssprache*. Leipzig: VEB Bibliographisches Institut.
- Frank, K.** (1992). *Sprachgewalt. Die sprachliche Reproduktion der Geschlechterhierarchie. Elemente einer feministischen Linguistik im Kontext sozialwissenschaftlicher Frauenforschung*. Tübingen: Niemeyer.
- 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.** (1991). *Untersuchungen zur Struktur der Lexik: Aufbau eines lexikalischen Basismodells*. Trier: WVT Wissenschaftlicher Verlag Trier.
- Herbermann, C.-P.** (1981). *Wort, Basis, Lexem und die Grenze zwischen Lexikon und Grammatik. Eine Untersuchung am Beispiel der Bildung komplexer Substantive*. München: Fink.
- Heringer, H.J.** (1984). Wortbildung: Sinn aus dem Chaos. *Deutsch als Fremdsprache*, 12, 1-13.
- Höhle, T.N.** (1985). On Composition and Derivation: The Constituent Structure of Secondary Words in German. In J. Toman (Hg.), *Studies of German Grammar* (S. 377-406), Dordrecht - Holland, Cinnaninson-USA: Foris Publications.
- Kastovsky, D.** (1990). The interaction of semantic and formal structures in the lexicon. In J. Tomasczyk & B. Lewandowska-Tomasczyk (Hg.), *Meaning and Lexicography*, Amsterdam, Philadelphia: Benjamins.
- Knop, S. de** (1987). *Metaphorische Komposita in Zeitungsüberschriften*. Tübingen: Niemeyer.
- Köhler, R.** (1986). *Zur linguistischen Synergetik: Struktur und Dynamik der Lexik*. Bochum: Brockmeyer.
- Köhler, R.** (1987a). Selbstregulation der Lexik. In H. Bluhme (Hg.), *Beiträge zur quantitativen Linguistik. Gedächtniskolloquium für Eberhard Zwirner, Antwerpen, 9.-12. April 1986* (S. 156-166), Tübingen: Narr.
- Köhler, R.** (1987b). System Theoretical Linguistics. *Theoretical Linguistics*, 14.2/3, 241-257.
- Köhler, R.** (1990a). Linguistische Analyseebenen. Hierarchisierung und Erklärung im Modell der sprachlichen Selbstregulation. In L. Hřebíček (Hg.), *Glottometrika 11* (S. 1-18), Bochum: Brockmeyer.
- Köhler, R.** (1990b). Synergetik und sprachliche Dynamik. In W.A. Koch (Hg.), *Natürlichkeit der Sprache und der Kultur* (S. 96-112), Bochum: Brockmeyer.

- Köhler, R.** (1991). Diversification of Coding Methods in Grammar. In U. Rothe, (Hg.), *Diversification Processes in Language: Grammar* (S. 47-55), Hagen: Margit Rottmann Medienverlag.
- Krylov, J.K.** (1982). Eine Untersuchung statistischer Gesetzmäßigkeiten auf der paradigmatischen Ebene der Lexik natürlicher Sprachen. In H. Guiter & M.V. Arapov (Hg.), *Studies on Zipf's Law* (S. 234-262), Bochum: Brockmeyer.
- Lewandowski, T.** (1990). *Linguistisches Wörterbuch 1-3*. Heidelberg, Wiesbaden: Quelle & Meyer.
- Ludewig, W.** (1991). Lexikon der Deutschen Sprachlehre. In G. Wahrig (Hg.), *Deutsches Wörterbuch* (S. 27-120), Gütersloh: Bertelsmann.
- Luukhainen, M.** (1990). Vom Satz zum Wort. Wortzusammensetzung als stilistisch-lexikalisches Element bei Heinrich Böll. *Zeitschrift für Germanistische Linguistik*, 18, 288-305.
- Ortner, H.** (1984). Neuere Literatur zur Wortbildung. *Deutsche Sprache*, 2, 141-158.
- Ortner, H., & Ortner, L.** (1984). *Zur Theorie und Praxis der Kompositaforschung. Mit einer ausführlichen Bibliographie*. Tübingen: Narr.
- Paprotté, W.** (1983). *Semantic Indeterminancies in Child Language, Part I*. Trier: LAUT (Linguistic Agency University of Trier), (Series A, Paper 80).
- Pieper, U.** (1979)., *Über die Aussagekraft statistischer Methoden für die linguistische Stilanalyse*. Tübingen: Narr.
- von Polenz, P.**, Wortbildung. In H.P. Althaus, H. Henne & H.E. Wiegand (Hg.), *Lexikon der germanistischen Linguistik* (S. 169-180), Tübingen: Narr.
- Raether, A.** (1987). *Trends der Wortbildung im gegenwärtigen Deutsch*. Manuscript, Göttingen.
- Raether, A., & Rothe, U.** (1991). Diversifikation der deutschen Komposita: 'Substantiv plus Substantiv'. In U. Rothe (Hg.), *Diversification Processes in Language: Grammar* (S. 85-91), Hagen: Margit Rottmann Medienverlag.
- Rosch, E.**, et al. (1976). Basic Objects in Natural Categories. *Cognitive Psychology*, 8, 382-439.
- Rothe, U.** (1988). Polylexy and Compounding. In K.P. Schulz (Hg.), *Glottometrika 9* (S. 121-134), Bochum: Brockmeyer.
- Sambor, J.** (1984). Menzerath's Law and the Polysemy of Words. In J. Boy & R. Köhler (Hg.), *Glottometrika 6* (S. 94-114), Bochum: Brockmeyer.
- Sambor, J.** (1989). Polnische Version des Projekts "Sprachliche Synergetik. Teil I. Quantitative Lexikologie". In R. Hammerl (Hg.), *Glottometrika 10* (S. 171-197), Bochum: Brockmeyer.
- Schierholz, S.J.** (1991). *Lexikologische Analysen zur Abstraktheit, Häufigkeit und Polysemie deutscher Substantive*. Tübingen: Niemeyer.

- Schonebohm, M.** (1979). *Wortbildung, Text und Pragmatik: am Beispiel der Teil-von-Relation im Bereich der deutschen Nominalkomposita*. Lund: Liber Läraomedel.
- Schröder, M.** (1978). Über textverflechtende Wortbildungselemente. *Deutsch als Fremdsprache*, 13, 85-93.
- Schröder, M.** (1979). Zu Beziehungen zwischen Wortbildung und Polysemie. *Deutsch als Fremdsprache*, 5, 286-291.
- Schröder, M.** (1986). Werden unsere Wörter immer länger? *Sprachpflege*, 35.5, 61-65.
- Shaw, J.H.** (1979). *Motivierte Komposita in der deutschen und englischen Gegenwartssprache*. Tübingen: Narr.
- Stepanowa, M. D., & Fleischer, W.** (1985). *Grundzüge der deutschen Wortbildung*. Leipzig: VEB Bibliographisches Institut.
- Steiner, P.** (1995). Effects of Polylexy on Compounding. *Journal of Quantitative Linguistics*, 2.2, 133-140.
- Tränkle, U.** (1980). *Mathematische und statistische Methoden für Studierende der Psychologie Biologie Medizin Pädagogik Soziologie*. Münster: Aschendorff.
- Vögeding, J.** (1981). *Das Halbsuffix "-frei". Eine Theorie der Wortbildung*. Tübingen: Narr.
- Warren, B.** (1990). The Importance of Combining Forms. In W. Dressler et al. (Hg.), *Contemporary Morphology* (S. 111-132), Berlin: de Gruyter.
- Wahrig, G.** (1986). *Deutsches Wörterbuch*. München, Mosaik.
- Walser, M.** (1980). *Ein fliehendes Pferd*. Frankfurt am Main: Suhrkamp.
- Weber, H.J.** (1974). *Mehrdeutige Wortformen im heutigen Deutsch. Studien zu ihrer grammatischen Beschreibung und lexikographischen Erfassung*. Tübingen: Niemeyer.
- Weber, N.** (1990). *Maschinelle Lexikographie und Wortbildungsstrukturen*. Tübingen: Niemeyer.
- Wildgen, W.** (1980). *Textuelle Bedingungen der Einführung und Verwendung adhoc gebildeter Komposita*. Trier: LAUT (Linguistic Agency University of Trier), (Series A, Paper 80).