



Brauchen wir neue Wörter ? Neologismen als Problem und Aufgabe von Lexikographie und Computerlinguistik

Lothar Lemnitzer

lothar@sfs.uni-tuebingen.de

Münster, 12. November 2004



„Mit rund 5000 neu aufgenommenen Wörtern, wie beispielsweise *Billigflieger*, *Dosenpfand*, *Fotohandy*, *Genmais*, *Ich-AG*, *LAN-Party*, *Minijob* und *Sars*, ist das Wörterverzeichnis auf den aktuellen Stand gebracht.“ (Duden Band 1, Rechtschreibung, 23. Auflage 2004, Vorwort)



„One of our major concerns has been completeness ... Huge numbers of new words and phrases, so essential for those keeping in touch with changes in ideas and technology, are included (CIDE, 1995, Foreword)



Standardwörterbücher und neue Wörter

In Neuauflagen von (Lerner)Wörterbüchern werden relativ viele neue Wörter verzeichnet, weil:

- bei diesen Wörtern die Normunsicherheit der BenutzerInnen besonders groß ist
- die Aufnahme vieler neuer Wörter ein Marketingargument für den Verlag ist



Normunsicherheiten

- Schreibt man *Carsharing*, *Car-Sharing* oder *Car-sharing*?
- Wie wird *E-Zine* ausgesprochen?
- Heißt es *die E-Mail* oder *das E-Mail*?
- Welche Genitivform ist richtig: *des Piercings* oder *des Piercing*?
- Wie heißt der Plural von *Flyer*: *die Flyer* oder *die Flyers*?



Normunsicherheiten 2

- Schreibt man 'up-talk' oder 'up talk' oder 'uptalk'?
- Kann man 'uptalk' auch als attributives Adjektiv verwenden?
- Schreibt man 'monetize eyeballs' oder 'monetize the eyeballs'?
- Was bedeutet diese Wendung? In welchem Zusammenhang wird sie verwendet?



Spezialisierte Sammlungen

Lexikographische Sammlungen von Neuwörtern des Deutschen:

- Gertrude Harlass / Heinz Vater: Zum aktuellen deutschen Wortschatz. GNV Tübingen 1974
- Alfred Heberth: Neue Wörter. Neologismen in der deutschen Sprache seit 1945. Wien 1977
- Lemnitzer / Ule: die Wortwarte. Wörter für heute und morgen. Tübingen (2000-2004; online)
- Dieter Herberg/Michael Kinne/Doris Steffens: Neuer Wortschatz. Neologismen der 90er Jahre im Deutschen. de Gruyter, 2004



Spezialisierte Sammlungen

Lexikographische Sammlungen von Neuwörtern des Deutschen:

- The Barnhart Dictionary of New English Since 1963. New York 1973 (Fortführung 1980)
- online: <http://www.highlands.com/Lexik/bdc.htm> (Barnhart Dictionary Companion)
- Burchfield: A Supplement to the Oxford English Dictionary. 4 vol., 1972-1988
- Fifty Years Among the New Words: A Dictionary of Neologisms, 1941-1991 (John Algeo, eine regelmäßige Kolumne in 'American Speech')



- Vorspann
- Was sind Neuwörter / Neologismen?
- Typologische Einordnung von Neologismenwörterbüchern
- Benutzung von Neologismenwörterbüchern
- Struktur und Informationsprogramm von Neologismenwörterbüchern
- Wie entstehen Neologismensammlungen (am Beispiel der Wortwarte)
- Ausblick



Was sind Neologismen?

„Ein Neologismus ist eine lexikalische Einheit bzw. eine Bedeutung, die in einem bestimmten Abschnitt der Sprachentwicklung in einer Kommunikationsgemeinschaft aufkommt, sich ausbreitet, als sprachliche Norm allgemein akzeptiert und in diesem Entwicklungsabschnitt von der Mehrheit der Sprachbenutzer über eine gewisse Zeit hin als neu empfunden wird“ (Herberg et al. 20004, S. XII)

Stichwörter: Usualisierung, Lexikalisierung, Integration



Was sind Neologismen?

Kandidaten für ein Neologismenwörterbuch

- sind Wörter, Wortteile oder Wortgruppen (*downloadbar, Cyber-, monetize eyeballs*)
- sind *Neuprägungen (Handy)* und *Neubedeutungen (Maus)*
- sind bisher in keinem Wörterbuch verzeichnet (= Ausschluss bekannter Wörter)
- werden von vielen Sprechern ausreichend häufig verwendet (= Ausschluss von Okkasionalismen)
- diese Sprecher sind Durchschnittssprecher (= Ausschluss von Fachsprachen)



Was wird ausgeschlossen?

Als Neologismen nicht berücksichtigt werden

- transparente Komposita und Derivationen
- übertragene Verwendungen bekannter Bedeutungen von Wörtern
- Okkasionalismen
- Wörter von geringem Interesse
- Modewörter

(nach: Barnhart and Barnhart: The dictionary of Neologisms)



Neologismenwörterbücher sind Wörterbücher mit pragmatisch beschränkter Lemmaauswahl. Ihr Informationsprogramm ähnelt dem allgemeinsprachlicher Gesamtwörterbücher.



Subtypen von Neologismensammlungen

- Kumulative Sammlungen mit relativ kurzen Publikationszyklen (z.B. Algeo, Wortwarte),
- Ergänzungslieferungen (Komplemente) zu großen allgemeinsprachlichen Gesamtwörterbüchern (z.B. Burchfield),
- eigenständige Neologismenwörterbücher (z.B. Neologismen der 90er Jahre).



Benutzung und Benutzergruppen

Als Benutzer eines spezialisierten Neologismenwörterbuchs kommen in Frage:

- am normgerechten Gebrauch von neuen Wörtern Interessierte (SprachenlerInnen, sprachlich Tätige);
- an der lexikographischen oder linguistischen Beschreibung Sprache und des Sprachwandels Interessierte (LexikographInnen, LinguistInnen);
- interessierte Laien, SprachkritikerInnen etc.



Benutzungssituationen

- Die Wahl eines Neologismenwörterbuchs, um eine lexikalische Wissenslücke zu schließen, setzt einen hohen Grad an linguistischer und (meta)lexikographischer Kompetenz voraus.
- Andere Benutzungssituationen (zu Forschungszwecken oder zur Unterhaltung) sind wahrscheinlicher.



Struktur eines Neologismenwörterbuchs

Wir betrachten:

- die Makrostruktur
- die Mikrostruktur bzw. das Informationsprogramm

von Neologismenwörterbüchern am Beispiel der Wortwarte.



Die Elemente der Makrostruktur der Wortwarte:

- Einleitung und Hintergrund,
- chronologisch geordnete Tageswortlisten (alphabetisch),
- sekundäre Zugriffsstrukturen (alphabetisch, thematisch)

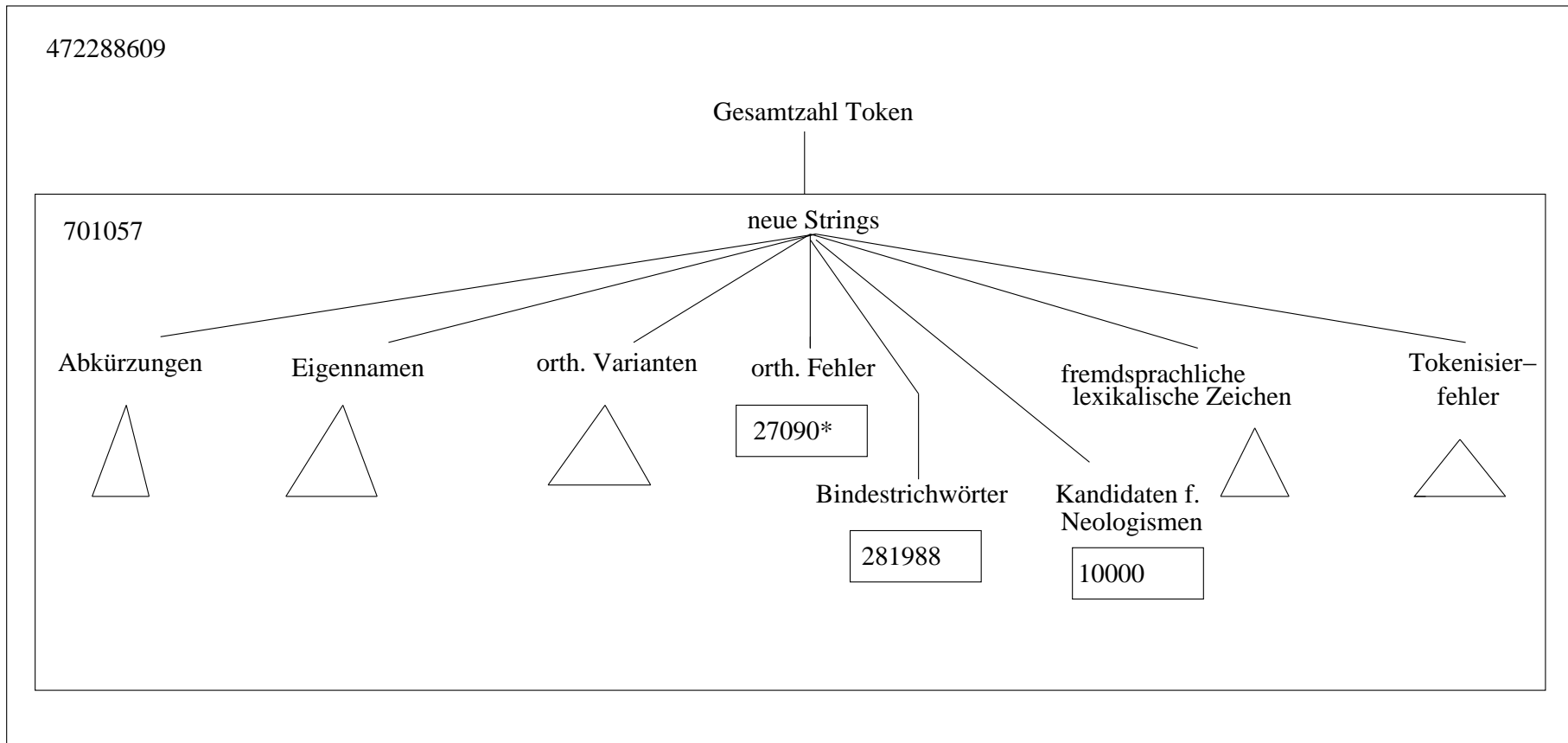


Mikrostruktur und Informationsprogramm

- Stichwort
- Wortart und Flexion
- Verwendungshäufigkeit (ermittelt über Google)
- Lexikalisches Feld
- Belegbeispiel



Mengengerüst der Rohdaten





Statistisches Profil der Wortwarte

Eintragskategorien und -zahlen in der WW
(Gesamt: 16700, 4. Nov. 2004)

- Adjektive: 952, Verben: 326,
- Informationstechnologie: 3137 Einträge (→ heise)
- Wirtschaft: 2344 Einträge (→ FTD)
- Technik, Politik, Gesellschaft: ca. 1100
- Telekommunikation, Technik, TV, Verkehr, Biotechnologie, Sprache, Medizin, Gesundheit, Sport, Bildung, Musik, Ernährung, Wissenschaft, Recht, Umwelt, Mode: < 1000 Einträge.



Wörterbucherstellung - am Beispiel der WW

Bei der Wörterbucherstellung können wir die folgenden Phasen unterscheiden

- Festlegung von Wörterbuchgegenstand und Wörterbuchbasis
- Datensammlung und -auswertung
- Lemmaselektion
- Erstellung von Einträgen
- Publikation



Wörterbuchgegenstand und -basis

- Gegenstand: Gegenwartsdeutsch, repräsentiert durch Zeitungstexte
- Auswahl von Tageszeitungen (Datenerhebung täglich)
- z.Z. ausgewählt: Handelsblatt, Süddeutsche, Financial Times, Rheinische Post, Tagesspiegel, Zeit, Spiegel, heise Website, Perlentaucher



Datensammlung und -auswertung

- Herunterladen der Texte auf den eigenen Rechner
- Bereinigung und Segmentierung der Texte
- Aufbau einer Wortliste (Index)
- Filter 1: Abgleich mit Referenzkorpus (80er und 90er Jahre)
- Filter 2: Abgleich mit Bestand der Wortwarte



Lemmaselektion

Die Lemmaselektion ist ein rein intellektueller Prozess. Ausgewählt werden:

- Relevante Neuprägungen
- Interessante Neuprägungen (aus linguistischer Sicht)
- Nicht-transparente Komposita und Derivationen
- Lehnwörter und Komposita mit Lehnwortbestandteilen



Nicht erfasst bzw. ausgewählt werden:

- Wortbestandteile (Affixe)
- Wortgruppen (wenn nicht durch Bindestrich gekoppelt)
- Neubedeutungen bereits registrierter lexikalischer Einheiten
- Transparente Komposita und Derivationen
- Rechtschreibfehler und Zeichenketten, die keine Wörter sind



Erstellung von Einträgen

Die ausgewählten Wörter werden

- mit einem Eintragsschema und
- mit Belegen aus dem Korpus verbunden

Weitere Aufgaben:

- Rückführung auf die Grundform (Ansetzungsform)
- Erstellung der Angaben (Wortart, Flexion, semantisches Feld)
- Auswahl und minimale Bearbeitung des Belegs



Ergänzung des Datenangebots

In unregelmäßigen Abständen

- werden die Lemmata mit Frequenzangaben versehen (über Google),
- werden die sekundären Zugriffsstrukturen (alphabetische Liste, Themenliste) erzeugt bzw. aktualisiert.



Die bearbeiteten Daten

1. werden in ein XML Format umgewandelt,
2. daraus wird eine HTML Datei erzeugt,
3. die HTML Datei wird als „neue Wörter“ auf den Server gestellt,
4. eine Kopie wird ins Archiv gestellt,
5. die Dateien für die Navigation im Archiv werden aktualisiert.



Rolle sprachtechnologischer Verfahren in der Neologismenlexikographie:

- Tokenisierung und Satzsegmentierung
- Lexikalische Statistik
- (geplant: Lemmatisierung und morphologische Analyse)



Nutzen der Aktualisierung lexikalischer Ressourcen für Sprach- und Texttechnologie

- Aktuellere Beschreibung von (neuen) Affixen
- Präzise lexikalische Kategorisierung, z.B. *simsen* = Kommunikationsverb, erlaubt Inferenzen
- Erkennung und Zusammenführung von Schreibvarianten
- Erkennung und Analyse von unknown words und named Entities, z.B. beim Wortartentagging
- Aus den Belegbeispielen: Muster für definitonische Kontexte.



Beispiel Affixe

- *Cyber-* - WW: 144/610; DEREKO: 1020/390
- *-[hH]opping* - WW: 1199/112; DEREKO: 3032/116

Ziel: valides statistisches Maß für die Differenz(en) finden (Produktivitätsmaß).



Beispiel 'unknown words'

- Allgemeines 'Gesetz' - Sehr viele Wörter kommen nur einmal in einem Textkorpus vor, unabhängig davon, wie groß das Textkorpus ist. D.h. dass sprachtechnologische Werkzeuge notwendigerweise mit 'Hapaxlegomena' konfrontiert sind.
- Eigene Beobachtung: der Anteil unbekannter Wörter neuer Texte, vor dem Hintergrund eines Textkorpus von 130 Mill. Token ist weit unter 1 %, bleibt aber nahezu konstant.



Fazit: Brauchen wir neue Wörter?

- Die TextproduzentInnen (ge)brauchen neue Wörter
- denn: Sprache ist ein generatives System
- also: ist es sinnvoll, diesem Phänomen Aufmerksamkeit zu schenken.



Vielen Dank
für Ihre Aufmerksamkeit!