

Burr, Elisabeth (2004): "Das Korpus romanischer Zeitungssprachen in Forschung und Lehre", in: Dahmen, Wolfgang / Holtus, Günter / Kramer, Johannes / Metzeltin, Michael / Schweickard, Wolfgang / Winkelmann, Otto (eds.): *Romanistik und neue Medien*. Romanistisches Kolloquium XVI (= TBL 455). Tübingen: Narr 133-162.

*Elisabeth Burr*

## **Das Korpus romanischer Zeitungssprachen in Forschung und Lehre**

### **1. Einleitung**

Nach Douglas Biber, Susan Conrad und Randi Reppen (1998: 1) läßt sich das Studium von Sprachen in zwei Hauptrichtungen einteilen, nämlich in "studies of structure and studies of use". Erstere Richtung entspricht der vor allem seit Chomsky traditionellen Linguistik des 20. Jahrhunderts, letztere z.B. der heutigen Korpuslinguistik. Die beiden Richtungen werden allerdings nicht als gleichwertig eingestuft, ebensowenig wie die beiden Entitäten *langue* und *parole* bzw. *competence* und *performance*, denen sie gewidmet sind. Im Gegenteil wird, wie nicht zuletzt auch Dell Hymes ausführte, die *parole* oder Performanz der *langue* bzw. Kompetenz gegenüber als minderwertig erachtet: „A major characteristic of modern linguistics has been that it takes structure as primary end in itself, and tends to depreciate use“ (Hymes 1972: 272).

Die traditionelle Beschäftigung mit der *langue* selbst, also mit der sprachlichen Struktur, charakterisieren nun Biber, Conrad und Reppen so:

Traditionally, linguistic analyses have emphasized structure – identifying the structural units and classes of a language (e.g., morphemes, words, phrases, grammatical classes) and describing how smaller units can be combined to form larger grammatical units (e.g., how words can be combined to form phrases, phrases can be combined to form clauses, etc.). (Biber/Conrad/Reppen 1998: 1).

Grundlage der dabei entwickelten Modelle ist, wie Michael Stubbs an in der modernen Linguistik einflußreichen Werken aufzeigen kann, eine verschwindend kleine Menge von sprachlichen Daten, die zudem in der Mehrzahl der Fälle aus erfundenen Sätzen besteht:

## 2 Elisabeth Burr

[...] it is so easy to be blind to the very small amount of data on which contemporary linguistics is based, and this is fundamental to the whole intellectual organisation of the discipline: the theories which linguists develop, the types of corroboration they claim, the methods they use, and the ways in which students are trained. (Stubbs 1993: 10).

Wallace Chafe, der sich mit den Methoden auseinandersetzt, die bei der Untersuchung von Sprache angewandt werden können, kommt zu einem ähnlichen Schluß „The techniques that have most dominated the field of linguistics [...] have been techniques focused on artificial rather than naturally occurring data.“ (Chafe 1992: 85). Dabei wird, wie das folgende Zitat von Pier Marco Bertinetto zeigt, das Erfinden von Beispielen sogar ausdrücklich der Untersuchung des Sprachgebrauchs vorgezogen:

La maggior parte degli enunciati illustrativi contenuti in questo lavoro sono frutto di libera invenzione; [...] non nego affatto che l'uso concreto offra formulazioni più ricche e più duttili di quelle che in genere sono partorite dalla mente di chi se le costruisce da sé secondo le esigenze del proprio argomentare; ma è indubbio che la speranza di trovare già bell'e confezionato l'esempio più adatto per ogni minima sfumatura di senso [...], appare un tantino ottimistica, data la marginalità di certi impieghi. (Bertinetto 1986: 13).

Dieses Vorgehen hat natürlich Auswirkungen auf die Theoriebildung. Wie diese abläuft, wird von Maurice Gross folgendermaßen beschrieben:

Currently, rules or syntactic theories are being elaborated for a small number of examples of a given phenomenon, and then the occasional discovery of new examples and counter-examples functions either to reinforce or to invalidate theories. (Gross 1994: 215).

Die Linguistik, die so bei der Theoriebildung verfährt, hat Charles J. Fillmore als Lehnstuhl-Linguistik bezeichnet:

He [the linguist] sits in a deep soft comfortable armchair, with his eyes closed and his hands clasped behind his head. Once in a while he opens his eyes, sits up abruptly shouting, „Wow, what a neat fact!“, grabs his pencil, and writes something down. Then he paces around for a few hours in the excitement of having come still closer to knowing what language is really like. (Fillmore 1992: 35).

## 2. Die Korpuslinguistik

Der Sprache im Sinne von Struktur oder *competence* setzt nun die Korpuslinguistik Sprache im Sinne von Sprachgebrauch und Sprechen gegenüber. Ich zitiere hierzu noch einmal Biber, Conrad und Reppen:

A different perspective [...] is to emphasize language use. From this perspective, we can investigate how speakers and writers exploit the resources of their language. Rather than looking at what is theoretically possible in a language, we study the actual language used in naturally occurring texts. (Biber/Conrad/Reppen 1998: 1).

Von einem solchen Perspektivenwechsel spricht auch Antoinette Renouf: „The aim of the linguist is to describe the language, and corpus linguistics reflects the shift in academic focus from the brain to the text as the appropriate source of information.“ (Renouf 1993a: 131).

Dieser Perspektivenwechsel scheint mir übrigens generell mit dem von Eugenio Coseriu schon 1955 geforderten Wechsel des Blickpunktes weg von einer Linguistik der Sprache zu einer Linguistik des Sprechens kongruent zu sein:

das Sprechen ist nicht von der Sprache her zu erklären, sondern umgekehrt die Sprache nur vom Sprechen. Das deswegen, weil Sprache konkret nur Sprechen, Tätigkeit ist und weil das Sprechen weiter als die Sprache reicht. Denn während die Sprache ganz im Sprechen steckt, geht das Sprechen nicht ganz in der Sprache auf. Daher muß unserer Meinung nach Saussures bekannte Forderung umgekehrt werden: statt auf den Boden der Sprache, muß man sich von Anfang an auf den des Sprechens stellen und dieses zur Norm aller anderen sprachlichen Dinge nehmen' (einschließlich der Sprache). (Coseriu 1955/1975: 258).

Den Begriff ‚Korpuslinguistik‘ gebraucht Coseriu in dieser Arbeit natürlich nicht.

## 2.1 Der Begriff *Korpuslinguistik*

*Korpuslinguistik* ist nämlich ein relativ neuer Begriff. Nach Geoffrey Leech hat er sich zwar schon 1984 mit dem Erscheinen des von Jan Aarts und Willem Meijs herausgegebenen Sammelbandes *Corpus linguistics* etabliert (cf. Aarts/Meijs 1984), unter Korpuslinguistik verstehen die sie Betreibenden aber nicht sofort eine neue Disziplin, sondern zunächst einmal eine spezifische Methodologie: „ ‚corpus linguistics‘ refers not to a domain of study, but rather to a methodological basis for pursuing linguistic research.“ (Leech 1992: 105), die prinzipiell in den unterschiedlichsten Bereichen der Linguistik zum Tragen kommen kann.

Obwohl diese Ansicht heute die wenigsten Korpuslinguistinnen und -linguisten teilen und obwohl sich schon seit Beginn der 90er Jahre hinter dem Begriff ‚Korpuslinguistik‘ ein sich als eigenständig verstehender Zweig der Linguistik verbirgt, wird die Korpuslinguistik z.T. auch weiterhin nicht als Disziplin, sondern als Methode beschrieben:

#### 4 Elisabeth Burr

Im Gegensatz zu den bisher vorgestellten Themenbereichen der Linguistik ist die Korpuslinguistik keine eigenständige Disziplin der Linguistik, sondern eine Methode, die die Disziplinen in korpus-basiert und nicht korpus-basiert unterteilt. (Stein 1998: 177).

Die Korpuslinguistik ist aber gerade nicht nur Untersuchung von Sprachen anhand von Korpora, sondern sie betrachtet, wie das folgende Zitat von Antoinette Renouf zeigt, auch das Sprechen anders, nämlich als zu einem großen Teil aus Mustern bestehend. Zudem versteht sie unter Korpora immer computerlesbare Korpora:

Corpus linguistics is a particular branch of linguistic study that is concerned with the analysis of patterns of words and word occurrences in a large, computer-held store of textual data. Such a store is referred to as a corpus. (Renouf 1993b: 173).

Die heutige Korpuslinguistik ist folglich ganz maßgeblich von den Möglichkeiten des Computers bestimmt und sollte sich nach Geoffrey Leech deshalb auch *computer corpus linguistics* (CCL) nennen, um sich von der ohne den Einsatz von Computern betriebenen Korpuslinguistik zu unterscheiden (cf. Leech 1992: 106).

#### 2.2 Der Begriff *Korpus*

Mit Korpora sind heute also Computer-Korpora gemeint, d.h. Zusammenstellungen von schon in elektronischer Form vorhandenen oder in elektronische Form überführten Texten. Bei den Texten handelt es sich um authentische Texte, die tatsächlich in einem wirklichen sozialen Kontext und ohne den Eingriff von Linguistinnen oder Linguisten schriftlich oder mündlich geäußert worden sind. Von Belegsammlungen unterscheiden sie sich dadurch, daß sie nicht nur die für ein bestimmtes sprachliches Phänomen im Sprechen gefundenen und damit differentiell gesammelten Belegstellen enthalten, sondern daß das mündliche oder schriftliche Sprechen als solches aufgenommen wurde.

Bei der Erstellung solcher Korpora wird zudem darauf geachtet, daß sie eine bestimmte Art des Sprechens abbilden und für sprachwissenschaftliche Untersuchungen geeignet sind:

A corpus is not simply a collection of texts. Rather, a corpus seeks to represent a language or some part of a language. The appropriate design for a corpus therefore depends upon what it is meant to represent. (Biber/Conrad/Reppen 1998: 246).

Gerade letzteres ist von Bedeutung, denn wie Habert, Nazarenko und Salem ausführen, hat die Verfügbarkeit elektronischer Texte in der Zwischenzeit

dazu geführt, daß der Begriff ‚Korpus‘ immer ungenauer wurde und im Bereich der Verarbeitung natürlicher Sprachen auch auf große Sammlungen von Texten angewandt wird, die rein aus opportunistischen Gründen zusammengestellt wurden:

Aujourd'hui que le texte électronique foisonne, des documents se trouvent parfois agrégés avant tout parce qu'ils sont faciles d'accès, sans que leur mise en relation ait été réellement pensée. (Habert/Nazarenko/Salem 1997: 144).

Aber auch eine dem Untersuchungsgegenstand entsprechende Zusammenstellung von elektronischen Texten ist noch kein Korpus im eigentlichen Sinne. Ein Korpus muß nämlich zumindest mit bibliographischen Annotationen angereichert sein, damit die einmal erhobenen Daten auch zugeordnet werden können. Zudem muß die Annotierung im Sinne der Austauschbarkeit und seiner Zusammenführung mit anderen Korpora heute bestimmten Standards entsprechen:

En effet, la simple existence sur support électronique ne fait pas d'un ensemble de textes un corpus électronique. Encore faut-il que ce document obéisse à des conventions de représentation, de codage répandues, voire consensuelles, qui permettent la transmission et la réutilisation des données textuelles en cause. (Habert/Nazarenko/Salem 1997: 145).<sup>1</sup>

### 2.3 Korpus-basierte Sprachbeschreibung

Die Korpuslinguistik geht also im Unterschied zur Systemlinguistik vom natürlichen Sprechen aus, das in Form von computerlesbaren, annotierten Korpora konserviert ist. Zudem eröffnet sie, da sie mit Hilfe des Computers die sprachlichen Realisierungen unter dem Blickwinkel von sogenannten *association patterns* (cf. Biber/Conrad/Reppen 1998: 4) untersucht, eine neue Sicht auf das Sprechen. Mit *association patterns* sind „the systematic ways“ gemeint, “in which linguistic features are used in association with other linguistic and non-linguistic features” (Biber/Conrad/Reppen 1998: 5), d.h. es geht um das regelmäßige Zusammenauftreten entweder bestimmter lexikalischer, grammatischer oder lautlicher Phänomene mit anderen lexikalischen, grammatischen oder lautlichen Phänomenen oder aber um ihren Gebrauch im Zusammenhang mit den verschiedenen Varietäten. Dieses

---

<sup>1</sup> Habert, Nazarenko und Salem meinen damit den von der *Text Encoding Initiative* (TEI) auf der Grundlage von SGML entwickelten internationalen Standard. Informationen zur TEI können unter <http://www.tei-c.org/> abgerufen werden. Dort finden sich auch die *Guidelines* (cf. ACH/ALC/ALLC 1994/1999).

Zusammenauftreten versucht die Korpuslinguistik dann auch zur Abgrenzung von Varietäten zu nutzen (cf. Biber/Conrad/Reppen 1998: 6). Dabei setzt sie nicht mehr nur quantitative Techniken ein, wie die frühe, sich aber nicht so nennende Korpuslinguistik.<sup>2</sup> Stattdessen kommt bei der Analyse der Daten heute sowohl die Intuition der Sprechenden einer Sprache und ihre Fähigkeit, die Daten zu interpretieren zum Tragen, als auch das Wissen, das Linguistinnen und Linguisten über die Sprache haben:

Recent corpus users have accepted that corpora, in supplying first-hand textual data, cannot be meaningfully analysed without the intuition and interpretative skills of the analyst, using knowledge of the language (*qua* native speaker or proficient non-native speaker) and knowledge *about* the language (*qua* linguist). In other words, corpus use is seen as a question of corpus *plus* intuition, rather than corpus *or* intuition. (Leech 1991: 74, zit. nach Partington 1998: 2).

### 3. Das Korpus romanischer Zeitungssprachen

Was nun das Korpus romanischer Zeitungssprachen betrifft, so habe ich es zunächst für meine eigenen Forschungen erstellt. Seine Zusammensetzung ergibt sich aus folgenden Überlegungen. Das Korpus sollte nicht die Zeitungssprache als solche repräsentieren, d.h. es sollte keinen irgendwie definierten sprachlichen Durchschnitt abbilden, den in Wirklichkeit niemand schreibt noch rezipiert.<sup>3</sup> Stattdessen sollte dieses Korpus einen Ausschnitt aus einem Sprechen so festhalten, wie dieses in der Wirklichkeit tatsächlich erscheint. Zugleich sollte dieses Sprechen aber auch ein Sprechen sein, das sich an viele richtet und das auch viele rezipieren.

Große Tageszeitungen sind ein solches Produkt. Zudem werden sie für ein bestimmtes Publikum produziert, zu dem es konkrete Informationen gibt. Da Zeitungen grundsätzlich auf Verbreitung aus sind, richten sie sich auch sprachlich an diesem Publikum aus. An ihrer Sprache können wir somit auch ablesen, welche Anforderungen sie meinen, an das sprachliche Wissen ihres Publikums stellen zu können.

Zudem sind Zeitungen selbst eine Einheit aus Varietäten und erfüllen demnach das besonders von Douglas Biber für Korpora aufgestellte Kriterium der Variabilität. Da diese Varietäten zusammen auftreten und sich die Bedeutung der in der Zeitung behandelten Themen an der ihnen zugestan-

---

<sup>2</sup> Gemeint ist hiermit die im Zusammenhang mit der Erforschung der Indianersprachen entstandene distributionalistische Korpuslinguistik.

<sup>3</sup> Mit den meisten Referenzkorpora wird gerade dieses Ziel verfolgt.

denen Textmenge zeigt, sollten die einzelnen Ausgaben auch genau so, wie sie in gedruckter oder elektronischer Form vorlagen, in das Korpus eingehen.

Als Erhebungszeitraum wurde jeweils ein wichtiger historischer Moment gewählt, weil sich damit eine gewisse thematische Einheitlichkeit verbindet und Zeitungen zu einem solchen Moment grundsätzlich eine größere Verbreitung erfahren. Der Gefahr, einen tagesspezifischen Sprachgebrauch zu konservieren, wurde mit der Berücksichtigung mehrerer Exemplare eines jeden Blattes begegnet.

### 3.1 Die Zusammensetzung des Korpus

Das Korpus romanischer Zeitungssprachen besteht aus zwei unterschiedlichen Komponenten und zwar einmal aus dem Korpus *Italienische Zeitungssprache* ‚*Deutsche Einigung 1989*‘ und dann aus dem Korpus *Französische, italienische und spanische Zeitungssprache* ‚*Europawahlen 1994*‘. Diese beiden Komponenten setzen sich selbst wieder wie folgt zusammen:<sup>4</sup>

---

<sup>4</sup> In ihrer Rohform besteht jede Komponente aus einer ganzen Wochenausgabe jeden Blattes. Die zweite Komponente enthält zudem auch noch weitere Blätter. Unter Korpus wird hier aber in Übereinstimmung mit dem unter 2.2 Ausgeführten nur der schon mit Annotationen angereicherte Teil verstanden.

8 Elisabeth Burr

Italienische Zeitungssprache ‚Deutsche Einigung 1989‘

<u>Zeitung</u>	<u>Quelle</u>	<u>Ausgabe</u>	<u>Wortformen/tokens</u>
<i>Corriere della Sera</i>	Zeitung	19.10.89	88.158
<i>Corriere della Sera</i>	Zeitung	20.10.89	79.030
<i>Corriere della Sera</i> <sup>5</sup>	Zeitung	21.10.89	91.099
Gesamt <i>Corriere</i>			258.287
<i>Il Mattino</i>	Zeitung	20.10.89	82.102
<i>Il Mattino</i>	Zeitung	21.10.89	89.399
Gesamt <i>Mattino</i>			171.501
<i>La Repubblica</i>	Zeitung	20.10.89	88.961
<i>La Repubblica</i>	Zeitung	21.10.89	85.997
Gesamt <i>Repubblica</i>			174.958
<i>La Stampa</i>	Zeitung	20.10.89	69.964
<i>La Stampa</i>	Zeitung	21.10.89	49.807
Gesamt <i>Stampa</i>			119.771
Gesamt Korpus			724.517

Französische, italienische und spanische Zeitungssprache ‚Europawahlen 1994‘

<u>Zeitung</u>	<u>Quelle</u>	<u>Ausgabe</u>	<u>Wortformen/tokens</u>
<i>Le Monde</i>	CD-ROM	12./13.06.94	89.786
<i>Le Monde</i>	CD-ROM	14.06.94	70.936
<i>Le Monde</i>	CD-ROM	15.06.94	75.514
Gesamt <i>Le Monde</i>			236.236
<i>Corriere della Sera</i>	CD-ROM	13.06.94	102.976
<i>Corriere della Sera</i>	CD-ROM	14.06.94	102.441
<i>Corriere della Sera</i>	CD-ROM	15.06.94	98.224
Gesamt <i>Corriere</i>			303.641
<i>La Vanguardia</i>	Magnetband	13.06.94	86.468
<i>La Vanguardia</i>	Magnetband	14.06.94	94.251
<i>La Vanguardia</i>	Magnetband	15.06.94	80.414
Gesamt <i>Vanguardia</i>			261.133

Fig. 1: Zusammensetzung des *Korpus Romanische Zeitungssprachen*

<sup>5</sup> Diese Ausgabe kam erst nach Abschluß der Untersuchung hinzu, für die das Korpus ursprünglich erstellt wurde.

### 3.2 Die Annotierung

Beim *Korpus romanischer Zeitungssprachen* handelt es sich, wie oben schon angedeutet, um ein annotiertes Korpus. Es wurde also zum einen im Sinne der beabsichtigten Analysen, zum anderen im Sinne der Austauschbarkeit mit bestimmten Informationen angereichert.

#### 3.2.1 Der Header

Die einzelnen Zeitungsexemplare wurden zunächst einmal mit einem sogenannten *Header* versehen, der wie Fig. 2 zu entnehmen ist, Informationen zur Herkunft der Texte, zu ihrer Form, zur Behandlung bestimmter sprachlicher Formen, zur Disambiguierung von Zeichen<sup>6</sup> und zur Form des Markups enthält. Das Englische erscheint hier im Sinne einer *lingua franca*.

```

<<!--French Newspaper Corpus "European Elections 1994" Component 3      -->
<<!--Original text copyright Le Monde, Paris                             -->
<<!--Text source Le Monde sur CD-ROM 1994                               -->
<<!--Copyright Research Publications Ltd                                 -->
<<!--CD-ROM Owner Elisabeth Burr                                       -->
<<!--Format MS-DOS Text                                                -->
<<!--not lemmatised                                                    -->
<<!--composed prepositions like "du", enclitici as in "admet-on" are retained as one form -->
<<!--word combinations with dash are retained as one form              -->
<<!--Accents combined with letters                                     -->
<<!--Apostrophee in its original function is followed by blank, i.e. "qu' elle" -->
<<!--Where hyphenation, line breaks correspond to printed version      -->
<<!--Where no hyphenation, line breaks do not correspond to printed version -->
<<!--Disambiguation: punctuation and digits, i.e. # = decimal point, \ = decimal comma -->
<<!--Mistakes not corrected, but annotated {sic}                       -->
<<!--Markup textual features                                           -->
<<!--Markup format COCOA                                              -->
<<!--Values of variables given in Italian                              -->
<<!--Copyright Elisabeth Burr                                          -->
<<!--The June 15, 1994 issue of the newspaper Le Monde, Paris, France -->

```

Fig. 2. Der Header des französischen Subkorpus

---

<sup>6</sup> Disambiguiert wurden Zeichen, die wie z.B. der Punkt (Satzzeichen, Dezimalpunkt), das Komma (Satzzeichen, numerisches Zeichen) oder die Anführungszeichen, die auch als Abkürzung für Minute bzw. Sekunde erscheinen, eine doppelte Funktion realisieren.

### 3.2.2 Die Kodierung nicht-sprachlicher Einheiten

Die Korpora wurden dann mit der in Fig. 3 aufgeführten Kodierung angereichert, die die vom Medium Zeitung vorgegebenen nicht-sprachlichen Einheiten, wie Sparten, Textarten etc. explizit macht.

Da dieses Markup-System im sogenannten COCOA-Format<sup>7</sup> im Großen und Ganzen im Rahmen der Erstellung des ersten italienischen Zeitungskorpus entwickelt wurde, tragen die verschiedenen Variablen bei der französischen und spanischen Komponente des Korpus bislang italienische Bezeichnungen.

---

<sup>7</sup> Das COCOA-Format ist zwar ein weit verbreitetes Format, in der Zwischenzeit gibt es aber, wie in Fn. 1 angemerkt, den von der TEI entwickelten internationalen Standard, so daß es notwendig erscheint, sobald Mittel und Wege gefunden sind, die Kodierung in diesen Standard zu überführen. Außerdem ist im Sinne einer Ausweitung der Analysemöglichkeiten eine viel weitreichendere Anreicherung mit Markup geplant. Dabei werden auch die von der *Expert Advisory Group on Language Engineering Standards* (EAGLES) erstellten *Guidelines* berücksichtigt werden.

## Korpus romanischer Zeitungssprachen 11

Variable	Kodierung	Beispiel
Zeitung als Fragment des Korpus	<Z>	<Z Stampa>
Ausgabe	<E>	<E 211089>
Sparte	<S>	<S Politica>
Positionierung des Textes	<C>	<C MEA01>
AutorIn	<N>	<N Ferrara Anna>
Sprache	<L>	<L Inglese>
Autorenschaft	<A>	
signiert		<A firmato>
anonym		<A Non firmato>
Redaktion		<A Redazione>
Texttype	<T>	
Vorzeile		<T Occhiello>
Schlagzeile		<T Titolo>
Untertitel		<T Sottotitolo>
Zusammenfassung		<T Sommario>
Zwischenüberschrift		<T Catenaccio>
Ankündigung		<T Civetta>
Artikel		<T Articolo>
'Aufmacher'		<T Spalla>
Leitartikel		<T Fondo>
Nachricht		<T Notizia>
Glosse		<T Corsivo>
Interview		<T Intervista>
Kolumne		<T Rubrica>
Kritik		<T Critica>
Kurzmeldung		<T Flash>
Kurznachricht		<T Breve>
Fernseh-, Kinoprogramm		<T Programma>
Filminhalt		<T Film>
Leserbrief		<T Lettera>
Liste		<T Elenco>
Wetterbericht		<T Tempo>
Buch-, Film-, Liedtitel, etc.		<T Nome>
Bildunterschrift		<T Foto>
Art des Sprechens	<P>	
fortlaufender Text		<P Prosa>
Zitat von schriftlicher Quelle		<P Citazione>
Zitat von mündlicher Quelle		<P Discorso>
Frage im Interview		<P Domanda>
Antwort im Interview		<P Risposta>

Fig. 3: Das Markup-System des *Korpus romanischer Zeitungssprachen*

### 3.2.3 Die Kodierung sprachlicher Einheiten

Was die Kodierung der sprachlichen Einheiten betrifft, so unterscheiden sich hier die frühere und spätere Komponente des Korpus ganz wesentlich. Dies hat damit zu tun, daß sie ursprünglich für unterschiedliche Unter-

## 12 Elisabeth Burr

suchungen erstellt wurden. Während nämlich beim italienischen Korpus zur Zeitungssprache 1989 alle finiten Verbformen mit einem Kode aus Lettern und Zahlen versehen wurden (cf. Fig. 4), fehlt eine solche Kodierung im späteren Korpus zu mehreren romanischen Zeitungssprachen.

Variable	Kodierung
finite Verbform	\$
Modus	
Indicativo	I
Congiuntivo	C
Imperativo	B
Tempus	
Presente	110
Passato prossimo (Passato composto)	111
Passato remoto (Passato semplice)	120
Trapassato remoto (Trapassato semplice)	121
Futuro semplice	130
Futuro anteriore (Futuro composto)	131
Imperfetto	210
Trapassato prossimo (Trapassato composto)	211
Condizionale presente (Condizionale semplice)	230
Condizionale passato (Condizionale composto)	231
Diathese	
Attivo	a
Passivo	p
Aspekt	
nicht Verbalperiphrase	00
stare + Gerundio	01
[...]	[...]
stare + cominciare a + Infinitivo	24
Person	
1. Person	1
2. Person	2
3. Person	3 <sup>8</sup>

Fig. 4: Die Kodierung der finiten Verben

Der den finiten Verbformen (\$) zugeordnete Kode setzt sich demnach zusammen aus den entsprechenden Werten der Variablen *Modus*, *Tempus*, *Diathese*, *Aspekt* und *Person* und macht insgesamt neun Zeichen aus. Für *ha mangiato* z.B. ist der Kode \$I111a003.

---

<sup>8</sup> Der Numerus wurde nicht berücksichtigt.

### 3.3 Das Korpus in der Forschung

Im folgenden soll nun anhand meiner eigenen Forschungen gezeigt werden, welche Art von Untersuchungen auf der Grundlage des hier beschriebenen Korpus und seines Markup unternommen werden können.

#### 3.3.1 Die Variation

Unter zur Grundlegung des von Coseriu vorgeschlagenen romanischen Verbalsystems sowie der Zahl der finiten Verben wurde die Variation im Innern des Korpus «Italienische Zeitungssprache 1989» untersucht (cf. Burr 1993). Besonderer Bedeutung kam dabei neben der Kodierung der sprachlichen Einheiten gerade der Kodierung der nicht-sprachlichen, vom Medium Zeitung selbst vorgegebenen Einheiten zu, die ich unter 3.2.2 vorgestellt habe. Überprüft wurde im Rahmen der Untersuchung die Signifikanz der Distribution der im Folgenden aufgeführten Kategorien:

Aktiv, Passiv, Verbalperiphrasen  
Indikativ, Konjunktiv, Imperativ  
Konjunktiv: aktuelle Ebene versus inaktuelle Ebene  
Konjunktiv aktuelle Ebene: Tempora  
Konjunktiv inaktuelle Ebene: Tempora  
Indikativ: aktuelle Ebene versus inaktuelle Ebene  
Indikativ aktuelle Ebene: Tempora  
Indikativ inaktuelle Ebene: Tempora

Erhoben wurden die Daten mit Micro-OCP, einer PC-Version des *Oxford Concordance Programs*. Diese Erhebung ergab 48.390 finite Verbformen, die etwa 7,5% der Gesamtwortzahl entsprechen. Ihre Untersuchung erbrachte eine ganze Reihe von sehr interessanten und z.T. unerwarteten Resultaten, die ich hier kurz zusammenfassen will.

Was die innere Variation des Korpus als solchem betrifft, so folgt der Gebrauch der finiten Verben in *Il Mattino* und *La Stampa* insgesamt der Norm des Korpus selbst, der *Corriere* weicht davon aufgrund eines geringeren Gebrauchs des Imperativs und der Tempora der inaktuellen Ebene des Indikativs ab. *La Repubblica* kommt aufgrund der vielen positiven Abweichungen eine ausdrückliche Sonderstellung zu.

Die synphasischen Einheiten insgesamt differenzieren sich v.a. aufgrund der Realisierungen der Tempora des Indikativs, zudem betreffen die signifikanten Abweichungen von der jeweiligen Norm fast durchweg die inaktuelle Ebene bzw. sind von ihr verursacht. Modus, Diathese sowie die Zahl der

finiten Verben spielen dagegen eine untergeordnete Rolle und die Gesamtzahl der aspektiven Verbalperiphrasen ist nur ausnahmsweise signifikant.

Die Titulatur, die sich durch ein reduziertes Verbalsystem auszeichnet,<sup>9</sup> erweist sich als eine relativ homogene Varietät. Ihr Verbalsystem wird aber unerwarteterweise allein von der Schlagzeile vollständig realisiert. Dabei lassen sich zwei unterschiedliche Normen ausmachen. So gebrauchen der *Mattino* und *La Stampa* in der Schlagzeile signifikant weniger *passato prossimo*, der *Corriere* und *La Repubblica* dagegen signifikant mehr. *Occhiello*, *Sottotitolo* und *Sommario* sind dagegen als zeitungübergreifende Texttraditionen zu interpretieren.

Besonders interessant sind die die Arten des Sprechens betreffenden Resultate, weisen sie doch darauf hin, daß in den Zitaten von schriftlichen Quellen (*Citazione*) ein schreibsprachlicher Stil zum Ausdruck kommt<sup>10</sup> und die Zeitungen beim Zitieren mündlicher Quellen (*Discorso*) eine Nachahmung des Gesprochenen anvisieren,<sup>11</sup> das sich einerseits durch einen signifikant stärkeren Gebrauch finiter Verben, des Präsens, Imperativs, Konjunktivs, Imperfekts und des Konditional Präsens und andererseits durch einen signifikant schwächeren des *passato prossimo*, des Passivs, Indikativs, Plusquamperfekts und des Konjunktivs Perfekt charakterisiert. Unerwartet sind die beiden auch hier wieder feststellbaren unterschiedlichen Normen. Innerhalb des *Discorso* realisieren nämlich der *Corriere* und *La Repubblica* zusammen signifikant weniger *passato remoto*, der *Mattino* und *La Stampa* realisieren dagegen signifikant mehr, obwohl sie aus diametral entgegengesetzten sprachlichen Regionen stammen. Das Ergebnis läßt sich interpretieren, wenn wir den Gebrauch des *passato remoto* in den anderen Einheiten der beiden Zeitungen betrachten. Beim *Mattino* ist nämlich nur der *Discorso* durch einen signifikant starken Gebrauch des *passato remoto* charakterisiert, alle als schriftsprachlich einzuordnenden Varietäten (*Citazione*, *Articolo*, *Notizia*) weisen dagegen gerade einen Mangel an *passato remoto* auf. Bei der Stilisierung des Gesprochenen scheint sich der *Mattino* also am neapolitanischen *italiano regionale* zu orientieren, wo das *passato remoto* das Vergangenheitstempus schlechthin ist. *La Stampa* dagegen erweist sich als eine insgesamt durch einen starken Gebrauch des *passato remoto* charakterisierte Zeitung.

---

<sup>9</sup> Nicht realisiert werden das *futuro anteriore* und der Konjunktiv Imperfekt.

<sup>10</sup> Nominalstil, signifikant viel Passiv und *trapassato prossimo*, kein *futuro anteriore*.

<sup>11</sup> Dardano spricht diesbezüglich von "parlato stilizzato" (Dardano 1987: 64) oder auch von der *finta oralità* (cf. Dardano/Frenguelli/Lauta 2000).

Die Spartenvariation ist am größten bei *La Repubblica* und am geringsten bei *La Stampa*. *Il Mattino* und der *Corriere* weisen dagegen beide eine mittlere Variationsbreite auf. Bei zwei Zeitungen läßt sich zudem jeweils eine Norm ausmachen, die nicht nur die jeweilige Sparte, sondern die Zeitung insgesamt charakterisiert. Da es sich beim *Mattino* um die Sparte *Cultura*, bei *La Repubblica* um die Sparte *Politica* handelt, können wir schließen, daß sich die beiden Zeitungen insgesamt an recht unterschiedlichen Stilnormen orientieren.

### 3.3.2 Die Referenz

Auf der Grundlage des mit einem weiteren Exemplar des *Corriere della Sera* integrierten Korpus zur italienischen Zeitungssprache 1989 wurden dann zwei Untersuchungen zur Referenz auf Frauen und Männer im aktuellen Sprachgebrauch durchgeführt.

Eine der beiden Untersuchungen galt der Referenz auf Frauen und Männer anhand der 15.385 Berufs-, Funktions- und Gradbezeichnungen des Korpus (cf. Burr 1995). Die Frage war, inwieweit Frauen in den Zeitungen überhaupt eine Rolle spielen, ob sie mit maskulinen oder femininen Formen bezeichnet werden, ob es dabei Unterschiede zwischen den Arten von Aktivitäten gibt, welche Feminisierungsverfahren verwendet werden und ob bei den angeblich generischen Formen tatsächlich Frauen mitgedacht werden. Die Analyse ergab, daß 94,25% der Nomina Agentis auf Männer oder männlich gedachte Agenten und nur 5,75% auf Frauen referieren. Die Mehrheit der in den untersuchten Zeitungen berücksichtigten Frauen erscheint zudem in traditionellen Rollen.

Die andere Untersuchung galt den personenbezogenen Eigennamen (cf. Burr 1997). Die Frage war, ob der Gebrauch von Eigennamen, die ja an sich nichts über die Eigenschaften des Referenten oder der Referentin aussagen, denen positive oder negative Konnotationen abgesprochen werden und die keine generische Funktionen übernehmen können, sondern lediglich auf konkrete Personen weiblichen oder männlichen Geschlechts referieren, Ausdruck einer Ungleichbehandlung der Geschlechter sein kann. Im Unterschied zur vorhergehenden Untersuchung ist dabei zu beachten, daß es hierbei nicht um die Eigennamen selbst gehen kann, sondern um ihr Zusammenauftreten mit kontextuellen Phänomenen. Diese können grammatischer, lexikalischer oder semantischer Natur sein. Hier soll allerdings nur in Auszügen auf den den grammatischen Asymmetrien gewidmeten Teil der Untersuchung eingegangen werden.

Erhoben wurden nur Eigennamen, die sich auf natürliche Personen beziehen, Firmennamen (*la Barilla*), Namen von Fällen (*il caso Renzi*), Ei-

genannten in literarischen Werken (*La morte di Danton*) und Eigennamen im Plural (*i Neri*) wurden also aussortiert. Die Gesamtzahl der erhobenen Eigennamen betrug 13.612.

Wie die Erhebung ergibt, beträgt die anhand von Eigennamen faßbare Präsenz von Frauen und Männern in den italienischen Tageszeitungen 9,03% zu 90,97%. Natürlich sagt dieses Ergebnis selbst noch nichts über eine Ungleichbehandlung der Geschlechter anhand der Eigennamen aus. Anders sieht die Lage aus, wenn wir bei den Eigennamen zwischen Vorname und Familienname unterscheiden:<sup>12</sup>

Referenz auf	Mann		Frau		Gesamt	
	absolut	%	absolut	%	absolut	%
Familienname	6904	55,76	183	14,89	7087	52,06
Vorname	477	3,85	297	24,17	774	5,69
Vor- und Familienname	5002	40,39	749	60,94	5751	42,25
Gesamt	12383	100,00	1229	100,00	13612	100,00

Bei der Referenz auf Frauen überwiegen, wie die Tabelle zeigt, grundsätzlich die das Geschlecht indizierenden Formen 'Vorname' und 'Vorname + Familienname' (85,11%), während bei Männern der oberflächlich sexindifferente Gebrauch des Familiennamens in 55,76% der Fälle bevorzugt wird. Ein Signifikanztest ergibt, daß die positive Abweichung von der Norm bei den verschiedenen Realisierungen der Referenz auf Frauen viel stärker ist als die negative der Realisierungen der Referenz auf Männer. Bei Frauen wird also ausdrücklich den beiden geschlechtsindizierenden Typen von Eigennamen der Vorzug gegeben, während sie bei der Referenz auf Männer zugunsten des Familiennamens eher vermieden werden.

Wie ist es nun mit dem maskulinen Artikel, der ja eigentlich die Aufgabe hat, entsprechend seiner primären, sich aus der systematischen Opposition zum femininen Artikel ergebenden Funktion, das Geschlecht des Referenten zu indizieren? Wäre der Familienname als solcher tatsächlich neutral, dann müßte auch bei männlichen Referenten der bestimmte Artikel stehen. In unserem Korpus präsentiert sich die Lage jedoch wie folgt:

---

<sup>12</sup> Hierbei ist zu beachten, daß lediglich der Vorname über das Geschlecht einer Person Auskunft gibt.

Referenz auf	Mann		Frau		Gesamt	
	absolut	%	absolut	%	absolut	%
Eigenname + bestimmter Artikel						
Vor- und Familienname	0	0	0	0	0	0,00
Vorname	0	0	1	0,34	1	0,13
Familienname	51	0,74	62	33,88	113	1,59

Bei 33,88% der Referenzen auf Frauen erscheint also der bestimmte Artikel, jedoch nur bei 0,74% der Referenzen auf Männer. Zudem entsprechen die Realisierungen des auf Männer referierenden Familiennamens fast genau der Norm, die der auf Frauen referierenden weichen dagegen überaus signifikant von dieser Norm ab.

Der Familienname ohne Artikel ist somit keineswegs sexindifferent, sondern verweist auf einen männlichen Referenten. Aufgrund seiner Maskulinisierung ist der bestimmte Artikel von seiner eingangs genannten Aufgabe befreit und kann nun eine Markierung ausdrücken, der in unserem Korpus ausschließlich die Bedeutung *berühmt/berüchtigt* zukommt. Da der bestimmte Artikel bei männlichen Referenten nur dort gebraucht wird, wo er auf etwas Außergewöhnliches hinweist, haben zugleich auch die Referentinnen, bei deren Familienname der Artikel erscheint, Ausnahmecharakter.

### 3.3.3 Die Phraseologie

Die letzte Untersuchung, auf die ich eingehen möchte, betrifft die Phraseologie im weiteren Sinne. Sie wurde anhand des Korpus zur französischen, italienischen und spanischen Zeitungssprache durchgeführt. Während die meisten im Rahmen der romanischen Sprachen unternommenen Untersuchungen zur Phraseologie von Wörterbüchern oder Phraseologiesammlungen ausgehen, setzte sie sich zum Ziel, Phraseologismen in natürlichen Texten aufzufinden, um somit auch Informationen über die tatsächliche Bekanntheit von Phraseologismen, ihren Gebrauch im Sprechen allgemein oder in verschiedenen Varietäten sowie zu ihrer kreativen Verwendung zu bekommen (cf. Burr 1997).<sup>13</sup> Auch bei dieser Untersuchung spielt wieder das Zusammenauftreten von sprachlichen Einheiten im Sprechen eine Rolle, diesmal allerdings im Sinne des sogenannten *idiom princi-*

---

<sup>13</sup> Mit Blick auf die kreative Verwendung von Phraseologismen ist der Wert *Nome* der Variablen <T> von besonderer Bedeutung (cf. 3.2.2), da sich nur mit Hilfe einer Explizitmachung der tatsächlich als Titel eines Buches, Films oder Liedes erscheinenden Textstellen die zu Phraseologismen gewordenen Buch-, Film- oder Liedtitel wie „la nave va“ oder „seis personajes en busca de autor“ etc. davon unterscheiden lassen.

*ple*, das nach John Sinclair (1991: 115) seine allgemeinste Realisierung als *collocation* erfährt. Die folgenden Beispiele sollen einen kleinen Einblick in die vielfältigen Erkenntnisse geben, die eine von den Kollokationen ausgehende und anhand eines elektronischen Korpus durchgeführte Untersuchung ermöglichen kann.<sup>14</sup>

Das erste Beispiel entstammt der Untersuchung zu den aus anderen historischen Sprachen übernommenen Elementen. Sie ergibt, daß im italienischen Teilkorpus die meisten solcher Einheiten auftreten. Hier die Ergebnisse zu den lateinischen Elementen:

nel+l' uomo inelegante che, **absit iniuria**, nel+l' aspet  
 al+ l' altro. il presidente **ad interim** the'odore sindik  
 maniera indiscriminata, ma **ad personam**. quanto a pianos  
 igliato e anche minimalista **ante litteram**: madraperla e  
 due intellettuali berlinesi **ante muro**, dichiara guerra a  
 l+la precarietà'. l' ultimo **casus belli** riguarda il  
 e riuscire? "basterebbe un **casus belli**, chiude sibillin  
 numero uno de+l veneto. un **cursus honorum** che comincia  
 sprofondato sotto il 10. il **de profundis** per gli uomini  
 uzione a meta' del+le pene. **dulcis in fundo**, le pene non  
 il terzo e' stato assegnato **ex aequo** a+l russo andrea  
 katherine heigl si fa **ex aequo** desiderare e odiare  
 contestano la liquidazione **in bonis** del+le societa' in  
 ha fatto. poteva provar+ci **in extremis** convocando calci  
 creto su+l lavoro, rinviato **in extremis** da+ l governo,  
 - sembra esser+si risolta **in extremis** in favore de+i  
 a punti, sottratta a berzin **in extremis**. arrivo - (ultima  
 e mi faccio richiamare **in extremis**? una trama da riv  
 stendere l' ala protettrice **in partibus infidelium**; al+la  
 volta. quello che trasforma **ipso facto** eugenio montale  
 0. due minuscole gocce ne+l **mare magnum** del+l' opera fred  
 era chi andava a vedere l' **opera omnia** di roger corman!"  
 sotto il peso del+la sua **opera omnia**. e delfini non fu  
 +l '65 al+la luiss (allora **pro deo** insegnava nel+l'  
 de, vi si crede ciecamente, **quia absurdum**. chi ha un' opi  
 innamora di lei, con molti **qui pro quo**, che conducono il  
 onsolidar+si e arrivare a+l **redde rationem** con la vittori  
 l suo pensiero. b sic+come **scripta manent**, ecco quel che  
 i scalda quando dismentiamo **nostra vanitate**, trattando l'  
 televisiva. e, sic+come **verba volant**, ho scritto "se

Das italienische Teilkorpus zeichnet sich weiter durch die kreative Verwendung von aus anderen Sprachen übernommenen Einheiten aus:

---

<sup>14</sup> Erhoben wurden die Daten mit dem interaktiven Analyseprogramm TACT 2.1.

lagante. pensiero dilagante. **business is business**, tutto q  
den, il 19. lo so, quello **e' business**. quello e' anche arb  
nulla vietava la relazione. **no comment** da parte mia, pens  
svizzera, li spendero' la'. **yes comment**. calcio e politic

Besonders interessant sind auch in den einzelsprachlichen Komponenten  
des Korpus aufscheinende Ko-Okkurenzen wie die folgenden, an denen sich  
ablesen läßt, was an einem Phraseologismus fixiert ist und in welcher Form.  
Fixiert ist im vorliegenden Fall, wie aus den Belegstellen hervorgeht, *donn-*  
*un x coup de y*:

rtade. la ville lui a m'e'me **donne' un se'rieux coup de pouce**  
e l' inte'rieur a sans doute **donne' un bon coup de pouce**,  
firment que les iraniens ont **donne' un coup de pouce** a+les  
, par la m'e'me occasion, de **donner un coup de pouce** 'a l'  
epuis, tous les ans, elle **va donner un coup de main** pendant

Das letzte Beispiel betrifft Einheiten, die in der Form eines Vergleichs auf-  
treten und elativisch funktionieren. Wie die folgende Konkordanz zeigt,  
treten solche Einheiten auch im Französischen auf, obwohl hier der Elativ  
als solcher nicht existiert:

qui parfois leur **ressemblent comme de+les frères** -  
us faibles **se font retourner comme une cr'e'pe**. a+le  
premier temps: "Il est snob **comme une chasse a+le renard**  
ire visiblement à **être connu comme le loup blanc**. Et y  
arure corporelle **est vieille comme Adam ou Eve** et touche  
fillettes **a les joues roses comme de+les hortensias** que  
redescend, **le visage noirci comme un charbon** par l'

e i sudamericani **l'attendono come il Messia**.  
e cio+è da quando **sono sorti come funghi** negozi che presta  
un nuovo partito, **cresciuto come un fungo** fuori dal+la  
E saltella **instancabile come un folletto**, piglio da  
enica era **affranto e feroce, come un fuoco beffardo** ne+gli  
è un partito! **Tranquillo come un corpo**, come un  
lo urlava Berti Vogts, **rosso come un peperone**. Mancavano  
viene neutralizzato, **legato come un salame** con fil di  
vita che rimarrà **indelebile come un sogno bellissimo** ne+1

ley en el cuerpo, se quieren **como hermanos**. Uno de ellos c  
les de un color intensi'simo, **como de terciopelo**. Para  
En efecto, hemos Ì sufrido **como perros**. Pero Indurain  
acceso a+el Cercle atraí'dos **como Ulises por las sirenas**,  
El sen'or Joan Barril escribe **como los a'ngeles**, pero algun

### 3.4 Das Korpus in der Lehre

Unter Einsatz von Teilen des *Korpus romanischer Zeitungssprachen* führe ich seit 1990 Studierende der Romanistik an computergestützte Korpusuntersuchungen heran, und zwar nicht nur, um ihnen die entsprechenden Methoden zu vermitteln, sondern auch um ihnen Einblicke in das tatsächliche Funktionieren von Sprache im Sprechen zu ermöglichen.

Das Projekt selbst hat in der Zwischenzeit eine ganze Reihe von Phasen durchlaufen, in denen nicht nur unterschiedliche Datenerhebungs- und Analyseprogramme (*OCP*, *Logiciel St. Chef*, *DBT* oder *TACT 2.1*) zum Einsatz kamen, sondern auch verschiedene Veranstaltungsformen durchgespielt wurden. Als bisher sinnvollste Vorgehensweise hat sich die Integration von Korpusanalysen in traditionelle Linguistikseminare erwiesen. Möglich wurde eine solche Integration 1998 durch das Erscheinen von *TACTWeb*, der WWW-Version von *TACT 2.1*, das die Erhebung von Daten aus online-verfügbaren Korpora erlaubt:

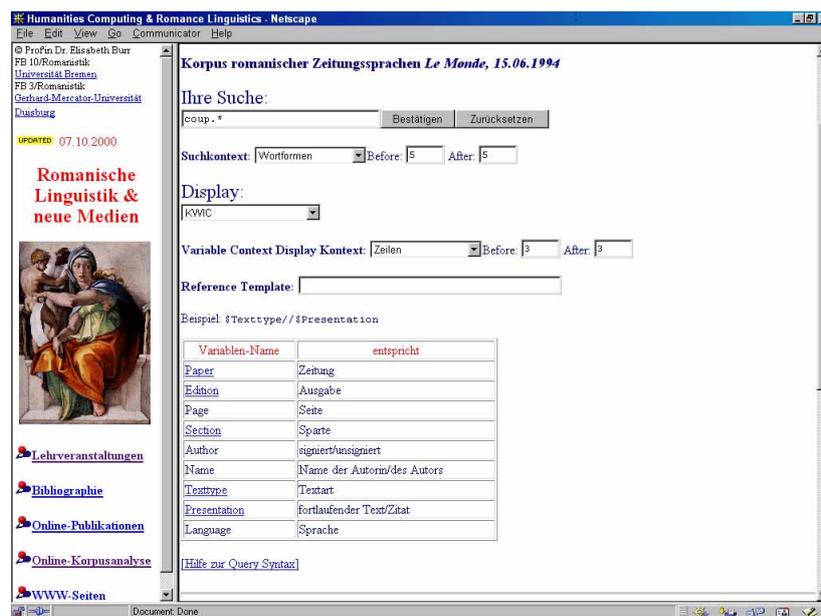


Fig. 5: Die Eingabemaske von TACTWeb

Online zur Verfügung stehen bisher die folgenden Komponenten:

Zeitung	Ausgabe	Wortformen
Corriere della Sera	21.10.89	91.099
Il Mattino	21.10.89	89.399
Corriere della Sera	15.06.94	98.224
Le Monde	15.06.94	75.514
La Vanguardia	15.06.94	80.414

Fig. 6: Die online-verfügbaren Komponenten des *Korpus romanischer Zeitungssprachen*

Die zu erhebenden Daten lassen sich von der Eingabemaske aus als Wortliste, KWIC-Konkordanz oder mit variablem Kontext anfordern. Bei allen Ausgabeformen ist das Schlüsselwort mit dem Text selbst verlinkt, so daß ein direkter Zugriff darauf möglich ist (cf. Fig. 7).

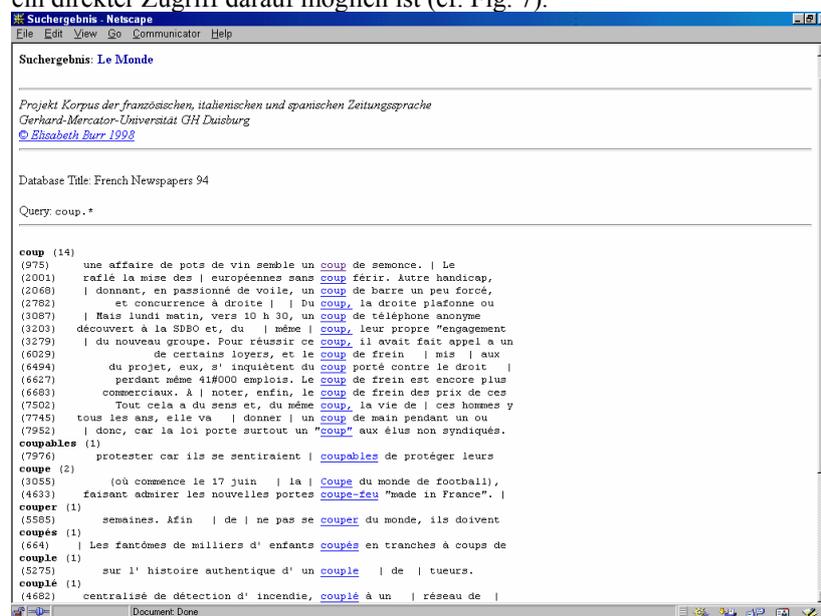


Fig. 7: Ergebnissenster einer Datenerhebung unter TACTWeb

Die im folgenden vorgestellten Untersuchungsergebnisse sollen einen Eindruck davon vermitteln, was Studierende in den unterschiedlichen Studienabschnitten in solchen Lehrveranstaltungen leisten.<sup>15</sup>

<sup>15</sup> Ich referiere im Folgenden aus den Seminararbeiten der Studierenden.

### 3.5 Tempus und Aspekt

Um eine Untersuchung zum Gebrauch von Tempus- und Aspektformen anhand der Teilkorpora zur französischen und spanischen Zeitungssprache durchführen zu können, mußten die Studierenden in Siegen die entsprechenden Formen zunächst einmal kodieren. Die getaggten Teilkorpora wurden jeweils wieder ins Netz gestellt, so daß dann mit TACTWeb Daten erhoben werden konnten (cf. Fig. 8).



Fig. 8: Die von Studierenden erarbeiteten Versionen des spanischen Teilkorpus

#### 3.5.1 *Perfecto compuesto* - *perfecto simple*

Eine Studentin im Grundstudium ging von der Sekundärliteratur zu *perfecto simple* und *perfecto compuesto* aus und untersuchte dann den Gebrauch dieser beiden Tempora in den Sparten *Política*, *Internacional* und *Deportes* von *La Vanguardia* (cf. Wahlers 1999). Dabei stellte sie unter anderem fest, daß, was die Zeitangaben betrifft, der Tempusgebrauch dem in der Sekundärliteratur Dargestellten entspricht, d.h. *perfecto simple* steht mit *pasado* (*el año pasado*), *hace* (*hace un mês*), *ayer*, *anoche*, *el 23 de marzo, en 1991*. Manchmal erscheint bei solchen Zeitangaben im Korpus auch das *imperfecto* oder *plusquamperfecto*, aber nie das *perfecto compuesto*, da diese Zeitan-

gaben das Ereignis in eine von der Gegenwart autonome Vergangenheit positionieren würden.

Beim *perfecto compuesto* erscheinen zunächst einmal weniger Zeitangaben. Es erscheint mit *a lo largo de la semana, por el momento, ahora* und mit Angaben, die *este (este año)* enthalten. Diese Angaben positionieren das Ereignis innerhalb des sogenannten *presente ampliado*: “Esta afirmación ha sido reiterada estos días en la capital tunesina”. Das *perfecto compuesto* erscheint nie mit Angaben, die *esa* enthalten: “Esa misma noche Mendoza se alojó en el mismo hotel”.

Bei *desde* erscheint *presente, perfecto compuesto, imperfecto* und *plusquamperfecto* aber kein *perfecto simple*: “Al menos 800 palestinos han sido asesinados desde el inicio de la ‘intifada’ en 1987”, “La RADS es miembro de la OUA desde 1984”. Bei *hasta* und *siempre* erscheinen dagegen sowohl *perfecto compuesto* und *perfecto simple*: “el Frente Polisario ha presentado hasta ahora un total de 24000 solicitudes”, “no declaró sus bienes y patrimonio en ningún momento hasta el 8 de diciembre de 1993”.

Im fortlaufenden Text steht häufiger das *perfecto simple* (77%) als das *perfecto compuesto* (23%), in der wörtlichen Rede dagegen öfter das *perfecto compuesto* (67%) als das *perfecto simple* (33%). Beim fortlaufenden Text läßt sich in den drei Sparten, *Política, Internacional* und *Deportes*, kein unterschiedlicher Tempusgebrauch feststellen. In der wörtlichen Rede tritt dagegen das *perfecto compuesto* häufiger in der Sparte *Política* als in der Sparte *Deportes* auf.<sup>16</sup> Erklärt wird dieses Ergebnis mit der zeitlichen und emotionalen Nähe der sprechenden Personen zu den gerade eben abgelaufenen Europawahlen.

### 3.5.2 *Venir de faire qc. - aller faire qc.*

Im Rahmen einer Untersuchung zu *venir de faire qc.* und *aller faire qc.* verglich die Studentin u.a. die Belege von *aller faire qc.* mit denen des *futur simple* (cf. Hummel 1999). Sie stellte dabei fest, daß das Subjekt bei *devoir* als Hauptverb einer Periphrase fast immer eine Person ist und die auszuführende Handlung so wirkt, als sei sie für die handelnde Person unangenehm oder schwierig: „M. Chirac va devoir choisir“, „Le président du RPR va donc devoir dire comment il compte débloquent le système de gouvernement de la société“. Dagegen ist das Subjekt bei *devoir* im *futur simple* oft keine Person und die inhaltlichen Aussagen wirken deutlich neutraler: „Un choix

---

<sup>16</sup> In der Sparte *Internacional* gibt es kaum direkte Rede.

qui devra s'accomplir“. Auch in anderen Fällen scheint, wie das folgende Beispiel zeigt, die Entscheidung für das *futur simple* oder die Verbalperiphrase vom Verhältnis zwischen dem Subjekt und der Handlung abzuhängen:

Selon M. Ballardur, ‚ceux qui plaident pour le protectionnisme oublient une vérité simple, l'Europe va croître au mieux de 3% l'an, alors que l'Asie connaîtra une croissance de 8% à 10%‘.

### 3.5.3 Imperfekt - *passato remoto*

Das folgende Beispiel betrifft eine interessante Beobachtung aus einer komplexen Untersuchung zum Imperfekt und *passato remoto* auf der Grundlage des Korpus *Italienische Zeitungssprache 1989* (cf. Steinert-Schmitz 1998), wo wie gesagt die finiten Verbformen kodiert erscheinen und deshalb anhand des Zahlenkodes direkt erhoben werden können. Die Studentin stellt fest:

- im neapolitanischen *Mattino* (21.10.89) findet das *passato remoto* seltener Verwendung (2%) als im Mailänder *Corriere della Sera* (21.10.89), wo 3,2% der finiten Verben im *passato remoto* erscheinen,
- im *Mattino* treten die Formen des *passato remoto* in der Regel einzeln auf, während es im *Corriere* immer wieder zu Ballungen von *passato remoto* im gleichen Artikel kommt,
- im *Mattino* steht 6 mal eine passivische Form im *passato remoto*, im *Corriere* dagegen nur 2 mal,
- im *Mattino* machen *fu/fummo/furono* etwa 10% der Formen des *passato remoto* aus, während es im *Corriere* nur etwa 4,5% sind,
- im *Mattino* erscheint sehr oft *fu/fummo/furono*, um den Hintergrund einer Nachricht abzustecken, im *Corriere* dagegen *era/erano*: „per non dire di quel Morse che fu la bandiera dei successi Ignis.“ (*Mattino*), „Morse, infondo, era soprattutto un gran tiratore.“ (*Corriere*), „Forse sarebbe cambiato niente ricorda Klimsmann anche se noi in entrambe le partite fummo troppo ingenui.“ (*Mattino*)

Ihre Hypothese ist, daß sich gerade hier diatopische Unterschiede in der Sprache der beiden Zeitungen zeigen.

### 3.6 Referenz auf Frauen und Männer

In einem Proseminar zur spanischen Zeitungssprache erhob eine Studentin anhand von Suffixen etwa 2.550 Berufsbezeichnungen (cf. Zander 1999). Bei ihrer Untersuchung kommt sie unter anderem zu dem Schluß:

- in *La Vanguardia* dominieren maskuline Berufsbezeichnungen,

- Berufsbezeichnungen bei Frauen erscheinen, zumindest im Singular, in ihrer femininen Form,
- stehen Berufsbezeichnungen im Plural, werden Frauen dagegen hinter einem generischen Maskulinum versteckt,
- das *género común*, Doppelformen und neutrale Umschreibungen werden nur z.T. angewendet,
- feminine Berufsbezeichnungen werden durch den Kontext ab-, maskuline werden dagegen aufgewertet.

Zu letzterem Punkt führt die Studentin aus: „Während der *principe* ohne große Umschweife *La Casa de España en Chicago* besucht, *conversò la Infanta animadamente, mientras tomaba un refresco*. An anderer Stelle bemerkt man ihren *excelente aspecto ... una chaqueta de color fucsia a juego con la blusa y los zapatos*. Ihrer Meinung nach wird hieran deutlich, daß ausschließlich der *principe* als Repräsentant des Königshauses gesehen wird, an der *infanta* dagegen lediglich die äußere Erscheinung interessiert.

### 3.7 Berichterstattung über den Krieg im Kosovo

Zum Schluß soll noch ein kleines Projekt vorgestellt werden, zu dem es im Rahmen eines Proseminars zur italienischen Zeitungssprache kam. Das Thema des Projekts war „Die Sprache der Kriegsberichterstattung“. Der konkrete Anlaß war der Krieg im Kosovo. Da die von mir erstellten Korpora kein entsprechendes Material enthielten, hat ein Student ein kleines Korpus aus Artikeln von online-verfügbaren italienischen Zeitungen erstellt und es auf der Grundlage des unter 3.2.2 vorgestellten Systems mit Markup angereichert (cf. Rizzo 1999). Das kleine Korpus von 3.513 Wortformen wurde dann ins Netz gestellt und von Studierenden untersucht.

Eine Studentin interessierte sich vor allem für die von der Sprache transportierte Ideologie. Um diese herauszufiltern, hat sie elf Wörter ausgewählt, deren Verwendung ihrer Meinung nach darauf Rückschlüsse erwarten ließ (cf. Gironi 1999). Diese Schlüsselwörter werden im Folgenden zusammen mit der Frequenz ihrer Belegstellen aufgeführt:

guerra	7
pace	4
bomb.*	15
morti	2
forza/e	17
uomini	16
donna/e	1
truppe	13
vittoria	1

## 26 Elisabeth Burr

sconfitto 1  
civili 2

Die Ergebnisse, die sie dabei erzielte, werde ich nun anhand ihrer eigenen Argumentationskette und den als Belegen aufgeführten Konkordanzen wiedergeben. Die Ausgabe der Referenzen auf jeder Konkordanzzeile zeigt dabei zugleich ein paar der Zuordnungsmöglichkeiten, die durch das Markup bereitgestellt werden.<sup>17</sup>

L'argomento è la guerra, ma di guerra non si parla mai direttamente con le parole che la designano, quasi come se si tentasse di mascherare la verità, di renderla più accettabile:

La Stampa/Titolo NORMAN MAILER E LA **GUERRA** | Norman  
La Repubblica/Articolo autentico **scenario di guerra**: da | u  
La Repubblica/Articolo è in **assetto di guerra**, avrebbe  
IlSole24Ore/Articolo previsti **prima della guerra**: lo hanno  
IlSole24Ore/Articolo **Il giorno 62 della guerra** | del Kos  
Il Messaggero/Occhiello | | | Tra **guerra** e diploma  
Il Messaggero/Articolo **alla fine della guerra**. E propri

Di guerra si parla in termini di numeri di uomini che la Nato spedisce in Kosovo, raramente in numeri di morti, sebbene la morte sia la prima delle conseguenze della guerra:

La Repubblica di circa cinquantamila | **uomini** da mandar  
La Repubblica avanguardia di quei 28.000 **uomini** | e donne  
La Stampa BRUXELLES | Fino a cinquantamila **uomini** dislocati  
La Stampa Wesley Clark. Cinquantamila **uomini** che, dice  
La Stampa scopo". Ma anche cinquantamila | **uomini** che non e  
La Stampa l'Alleanza prevedeva che 28 mila **uomini** | bastass  
La Stampa quasi il doppio degli **uomini**: alla Nat  
La Stampa ha già oggi circa 14 mila **uomini** che | fan  
La Stampa dichiara le sue disponibilità di **uomini** e mezzi r  
La Stampa 'apporto più alto, con circa 9000 **uomini**, mentre g  
La Repubblica e attaccato con duecento **uomini** i villagg  
IlSole24Ore orza di interposizione a 45mila **uomini** | Rallent  
IlSole24Ore jugoslavi 45MILA **UOMINI** | PER LA  
IlSole24Ore in dispiegamento di 45mila **uomini**, invece d  
Il Messaggero alla Nato l'aumento a 50.000 **uomini** della for  
Il Messaggero al minimo le perdite di **uomini** e mezzi N

Di guerra si parla soprattutto in termini dei bombardamenti della Nato e raramente in termini di vittime civili:

La Repubblica | | | Milosevic **bombarda** villagg  
La Repubblica | | | Ancora una notte di **bombardamenti**. I  
La Stampa piedi una regione martoriata dai **bombardamenti** e

---

<sup>17</sup> Die Arbeit wurde auf Italienisch verfaßt.

## Korpus romanischer Zeitungssprachen 27

La Stampa Nato insiste sul successo dei **bombardamenti**. A  
La Stampa le ragioni per sospendere i **bombardamenti** -  
La Stampa la sua strategia o sospenda i **bombardamenti**".  
IlSole24Ore militari dopo 60 giorni di **bombardamenti** al  
Il Messaggero è vicina. Ancora due mesi di **bombardamenti** a  
Il Messaggero Quanto alla decisione di **bombardare** ad al  
La Repubblica per tutta la giornata - hanno **bombardato** e att  
IlSole24Ore | Gli aerei della Nato hanno **bombardato** una d  
IlSole24Ore velivoli Nato | avevano già **bombardato** un'al  
La Repubblica morti MORINI - Piovono le **bombe** della Nato  
Il Messaggero Nato Ancora due mesi di **bombe** | e Milose  
Il Messaggero le unità colpite della **bombe** dell'Allea

Pace viene usato soltanto in relazione all'esercito della Nato:

La Stampa "Kfor plus", la **forza di pace** potenziata  
La Stampa utilizzati per una **missione di pace** e per nesso  
IlSole24Ore futura **forza internazionale di pace** per il Koso  
Il Messaggero 50.000 uomini della **forza di pace** destinata a

Forza/forze non viene usato per indicare gli eserciti di Milosevic:

La Repubblica/Articolo il momento - di una **forza** da impiega  
La Repubblica/Articolo inviare da subito la **forza** ai | confi  
La Repubblica/Articolo partecipazione alla **forza** anche di m  
La Repubblica/Articolo dell'eventuale **forza** di interpo  
La Stampa/Articolo "Kfor plus", la **forza** di pace |  
La Stampa/Articolo si riconosce che la **forza** avrà un |  
La Stampa/Articolo finora. Sarà una **forza** - spiega i  
IlSole24Ore/Titolo 1999) | | | | | La **forza** di interpo  
IlSole24Ore/Catenaccio UOMINI PER LA **FORZA** | DI INTER  
IlSole24Ore/Articolo I nuovi piani per la **forza** militare d  
IlSole24Ore/Articolo di profughi. La **forza** di interpo  
IlSole24Ore/Articolo esatta della **forza** e i contri  
IlSole24Ore/Articolo della futura **forza** internazio  
Il Messaggero/Articolo 50.000 uomini della **forza** di pace de  
La Repubblica/Articolo "conferenza sulle **forze**" che dovrà  
La Stampa/Articolo Usa che comanda le **forze** dell'Allea  
La Stampa/Articolo di generazione delle **forze**", cioè l'o

La presenza delle donne nelle truppe della Nato viene a malapena presa in considerazione come se fosse una svista il fatto che anche le donne vivono la guerra:

La Repubblica di quei 28.000 uomini e **donne** che dovb

## 4. Forschung und Lehre

Wie aus dem oben Dargestellten hervorgegangen sein sollte, erlauben selbst Korpora, die wie das hier vorgestellte *Korpus Romanischer Zeitungssprachen* im Vergleich zu den großen Referenzkorpora einen relativ kleinen

Umfang haben,<sup>18</sup> das natürliche Sprechen systematisch und als „verbundene Rede“ (cf. Humboldt 1827-1829/1963: 186) zu untersuchen und dabei wie Charles J. Fillmore es ausdrückt, zu Erkenntnissen zu gelangen, die anders nicht hätten gewonnen werden können: „every corpus that I’ve had a chance to examine, however small, has taught me facts that I couldn’t imagine finding out about in any other way.“ (Fillmore 1992: 35). Mit der Korpuslinguistik eröffnet sich zudem ein Weg, die Linguistik von ihrer einseitigen Theorielastigkeit und Modellorientiertheit auf den Boden der Realität, d.h. auf die sprachlichen Fakten selbst zu holen und das Sprechen als einzige faßbare Realität der Sprache (wieder) ins Zentrum des Interesses zu rücken.

Was erforschbar ist, hängt dabei natürlich ganz maßgeblich davon ab, welche sprachlichen und nicht-sprachlichen Einheiten eines Korpus mithilfe von Markup explizit gemacht wurden bzw. welche Einheiten mithilfe der Datenerhebungssyntax auch ohne Markup einigermaßen genau abgegrenzt werden können. So lassen sich auf der Grundlage des *Korpus romanischer Zeitungssprachen* in seinem derzeitigen Zustand alle Arten von Untersuchungen durchführen, die von bestimmten Wortformen ausgehen oder die bestimmte formale Regelmäßigkeiten nutzen können. Zudem ist es aufgrund des Markup möglich, bei der Erhebung der Belegstellen nach den von den Zeitungen vorgegebenen nicht-sprachlichen Einheiten zu differenzieren. Das Gleiche gilt für die Kategorien Tempus, Aspekt und Modus in der rein italienischen Komponente von 1989, nicht jedoch in dem mehreren romanischen Sprachen gewidmeten Subkorpus von 1994. Sprachvergleichende Untersuchungen zu diesen Kategorien sind so also nicht möglich. Sehr mühsam bleibt ferner die Beschäftigung mit den Eigennamen oder den *Nomina Agentis* sowie in gewisser Weise auch die mit den Kollokationen. Zudem lassen sich ohne eine Kodierung der sogenannten *parts of speech* eine ganze Reihe von grammatischen Untersuchungen computergestützt überhaupt nicht durchführen. Das Markup muß also nicht nur, wie schon angedeutet, an den TEI-Standard angepaßt, sondern es muß auch auf möglichst viele sprachliche Einheiten ausgedehnt werden. Daß sich mit Blick auf die damit möglich werdenden Korpusuntersuchungen eine solche Arbeit lohnt, sollten nicht nur die Ergebnisse meiner eigenen Forschungen, sondern auch die der studentischen Projekte gezeigt haben.

Was die Lehre betrifft, so kann ich aufgrund meiner Erfahrungen mit der Computerkorpus-gestützten Lehre nur dafür eintreten, daß Korpuslinguistinnen und -linguisten bei den Studierenden das Verständnis für Korpora und ihre Untersuchung fördern und die Arbeit an elektronischen Korpora,

---

<sup>18</sup> Das *British National Corpus* z.B. umfaßt 100 Millionen Wortformen.

wo immer möglich, in Seminare zur Linguistik der verschiedenen romanischen Sprachen einbeziehen. Die Studierenden erfahren so nämlich ganz konkret, daß Sprache uns als ‚verbundene Rede‘ entgegentritt und nicht in Einzelementen. Zudem weckt die Arbeit mit Korpora ausdrücklich das Interesse am Funktionieren der Sprache und begünstigt das Sich-Stellen von Fragen. Darüber hinaus werden die Studierenden durch systematische, kontrollierbare und wiederholbare Untersuchungen in die Lage versetzt, vorliegende Sprachbeschreibungen mit ihren Ergebnissen zu vergleichen und anhand der von ihnen selbst erhobenen Daten kritisch zu hinterfragen.

Dabei auf online-verfügbare Korpora, die mit einem interaktiven Programm wie TACTWeb über das WWW untersucht werden können, zurückzugreifen, scheint mir bislang die beste Lösung, da so nicht der Computer als solcher im Zentrum steht, sondern die Arbeit am Korpus. Außerdem ermöglicht die relativ benutzungsfreundliche Oberfläche den Studierenden, erst einmal mit Abfragen und den dabei erhobenen Daten zu spielen und anhand der z.T. ganz unerwarteten Ergebnisse zu lernen, ihre Suchen genauer zu definieren, Hypothesen aufzustellen und nach Beweisen zu suchen. Überdies erwerben sie auf diese Weise, neben den anderen Schlüsselqualifikationen, die solche Lehrveranstaltungen vermitteln, gleichzeitig auch die in vielen Betätigungsfeldern einsetzbaren Kenntnisse im bewußten und kritischen Umgang mit den neuen Medien und Technologien.

## Bibliographie

- Aarts, Jan/Meijs, Willem (eds.) (1984): *Corpus Linguistics. Recent Developments in the Use of Computer Corpora in English Language Research* (= Costerus, N.S. 45). Amsterdam: Rodopi.
- ACH/ALC/ALLC (eds.) (1994/1999): *Guidelines for Electronic Text Encoding and Interchange* (Edited by C. M. Sperberg-McQueen and Lou Burnard). Chicago/Oxford: Text Encoding Initiative (TEI): <<http://www.tei-c.org/>>. 10.10.2000.
- Bertinetto, Pier Marco (1986): *Tempo, aspetto e azione nel verbo italiano: Il sistema dell'indicativo*. Firenze: Accademia della Crusca.
- Biber, Douglas/Conrad, Susan/Reppen, Randi (1998): *Corpus linguistics. Investigating language structure and use*. Cambridge: Cambridge University Press.
- Burr, Elisabeth (1993): *Verb und Varietät. Ein Beitrag zur Bestimmung der sprachlichen Variation am Beispiel der italienischen Zeitungssprache* (= Romanische Texte und Studien 5). Hildesheim: Olms, 591 pp.
- Burr, Elisabeth (1995): "Agentivi e sessi in un corpus di giornali italiani", in: Marcato, Gianna (ed.): *Donna & Linguaggio. Convegno Internazionale di Studi, Sappada/Plodn (Belluno), 26-30 giugno 1995*. Padova: CLEUP 141-157.
- Burr, Elisabeth (1997): "Neutral oder stereotyp. Referenz auf Frauen und Männer in der italienischen Tagespresse", in: Dahmen, Wolfgang/Holtus, Günter/Kramer,

### 30 Elisabeth Burr

- Johannes/Metzeltin, Michael/Schweickard, Wolfgang/Winkelmann, Otto (eds.): *Sprache und Geschlecht in der Romania*. Romanistisches Kolloquium X (= TBL 417). Tübingen: Narr 133-179.
- Burr, Elisabeth (1997): *Wiederholte Rede und idiomatische Kompetenz*. Französisch, Italienisch, Spanisch. Habilitationsschrift, Gerhard-Mercator-Universität GH Duisburg, Fachbereich 3: Sprach- und Literaturwissenschaften (Manuskript 455 pp. - Erscheint als TBL, Gunter Narr).
- Chafe, Wallace (1992): „The importance of corpus linguistics to understanding the nature of language“, in: Svartvik, Jan (ed.): *Directions in Corpus Linguistics*. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991 (= Trends in Linguistics. Studies and Monographs 65). Berlin/New York: Mouton de Gruyter 79-97.
- Coseriu, Eugenio (1955/1975): „Determinierung und Umfeld“, in: Coseriu, Eugenio: *Sprachtheorie und allgemeine Sprachwissenschaft*. 5 Studien (= Internationale Bibliothek für Allgemeine Linguistik/International Library of General Linguistics 2). München: Fink 253-290.
- Dardano, Maurizio (1987): „Il linguaggio dei giornali“, in: Jacobelli, Jader (ed.): *Dove va la lingua italiana?* Roma/Bari: Laterza 58-65.
- Dardano, Maurizio/Frenguelli, Gianluca/Lauta, Gianluca (2000): „Parlato e finzione di parlato nella stampa di oggi“, in: Comitato scientifico della SILFI (ed.): *«Tradizione & Innovazione». La linguistica e filologia italiana alle soglie di un nuovo millennio*. Reader VI Convegno Internazionale della Società Internazionale di Linguistica e Filologia Italiana, Gerhard-Mercator-Universität Duisburg, 28.06-02.07.2000. Duisburg: Universität Duisburg 94-95.
- Expert Advisory Group on Language Engineering Standards (EAGLES)* (1996): *EAGLES Guidelines* <<http://www.ilc.pi.cnr.it/>>. 11.10.2000.
- Fillmore, Charles J. (1992): „‘Corpus linguistics’ or ‘Computer-aided armchair linguistics’“, in: Svartvik, Jan (ed.): *Directions in Corpus Linguistics*. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991 (= Trends in Linguistics. Studies and Monographs 65). Berlin/New York: Mouton de Gruyter 35-60.
- Gironi, Francesca (1999): *Berichterstattung über den Krieg im Kosovo*. Uni-Duisburg: PS-Arbeit.
- Gross, Maurice (1994): „Constructing Lexicon-Grammars“, in: Atkins, Sue B.T./Zampolli, Antonio (eds.): *Computational Approaches to the Lexicon*. Oxford: Oxford University Press.
- Habert, Benoît/Nazarenko, Adeline/Salem, André (1997): *Les linguistiques de corpus*. Paris: Armand Colin.
- Humboldt, Wilhelm von (1827-1829/1963): „Über die Verschiedenheiten des menschlichen Sprachbaus“, in: Humboldt, Wilhelm von: *Schriften zur Sprachphilosophie*, Werke in fünf Bänden III. Darmstadt: Wissenschaftliche Buchgesellschaft 144-367.
- Hummel, Carolin (1999): *Die Verbalperiphrase im Französischen am Beispiel von venir de faire qc. und aller faire qc.* Uni-Siegen: HS-Arbeit.

- Hymes, Dell H. (1972): „On Communicative Competence“, in: Pride, J. B./Holmes, Janet (eds.): *Sociolinguistics* (= Penguin Modern Linguistics Readings). Harmondsworth: Penguin 269-293.
- Leech, Geoffrey (1992): „Corpora and theories of linguistic performance“, in: Svartvik, Jan (ed.): *Directions in Corpus Linguistics*. Proceedings of the Nobel Symposium 82, Stockholm, 4-8 August 1991 (= Trends in Linguistics. Studies and Monographs 65): Berlin/New York: Mouton de Gruyter 105-122.
- Partington, Alan (1998): *Patterns and Meanings*. Using Corpora for English Language Research and Teaching (= Studies in Corpus Linguistics 2). Amsterdam/Philadelphia: John Benjamins.
- Renouf, Antoinette (1993a): "Sticking to the text: a corpus linguist's view of language", in: *Aslib Proceedings* 45 (5): 131-136.
- Renouf, Antoinette (1993b): "What the linguist has to say to the information scientist", in: *The Journal of Document and Text Management* 1/2: 173-190.
- Rizzo Angelo (1999): *Die Sprache der Kriegsberichterstattung*. Uni-Duisburg: PS-Arbeit.
- Sinclair, John (1991): *Corpus, Concordance, Collocation* (= Describing English Language). Oxford: Oxford University Press.
- Stein, Peter (1998): *Einführung in die französische Sprachwissenschaft* (= Sammlung Metzler 307). Stuttgart/Weimar: Metzler.
- Steiner-Schmitz, Sabina (1999): *Der Verwendungsbereich von passato remoto - imperfetto in einem elektronischen Korpus*. Uni-Duisburg: HS-Arbeit.
- Stubbs, Michael (1993): „British Traditions in Text Analysis. From Firth to Sinclair“, in: Baker, Mona/Francis, Gill/Tognini-Bonelli, Elena (eds.): *Text and Technology*. In Honour of John Sinclair. Philadelphia/Amsterdam: John Benjamins 1-33.
- Wahlers, Carola (1999): *Perfecto simple und Perfecto compuesto im Spanischen*. Uni-Siegen: PS-Arbeit.
- Zander, Andrea (1999): *Feminine Berufsbezeichnungen in der spanischen Zeitung*. Uni-Siegen: PS-Arbeit.