

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

## **Das Korpus romanischer Zeitungssprachen und sein online-Einsatz in der Lehre**

**Elisabeth Burr**

### **1 Einleitung**

Die Linguistik dieses Jahrhunderts basiert insgesamt gesehen auf einer grundsätzlich dualistischen Sichtweise von Sprache. Dieser Dualismus betrifft nicht nur die funktionellen Oppositionen, die zumeist als binäre Oppositionen gesehen werden, sondern ganz besonders die Sprache selbst. Die vor allem seit Saussure übliche Unterscheidung zwischen *langue* und *parole* liegt auch heute noch den meisten Ansätzen zugrunde. Unterstützung erfährt sie nicht zuletzt durch Chomsky's Unterscheidung zwischen *competence* und *performance*.

Die beiden Pole werden allerdings nicht als gleichwertig erachtet. Im Gegenteil, im Zentrum des Interesses steht vor allem die *langue* bzw. Kompetenz. Ihrer Erfassung und Beschreibung gelten die entwickelten Modelle und die um sie geführte Diskussion. Dell Hymes beurteilt dies so: „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). Grundlage dieser 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:

[...] 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) und Maurice Gross beschreibt die gängige Vorgehensweise wie folgt:

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).

Fragen der empirischen Evidenz - und im Sinne von Roger G. van de Velde auch der empirischen Adäquatheit dieser Modelle werden nicht nur auf die Regionen der Performanz,

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

sondern oft sogar in den Aufgabenbereich anderer Wissenschaften geschoben. Eine Theorie der Performanz wird erst gar nicht entwickelt (cf. Velde 1979: 24) und noch weniger eine Theorie des Sprechens insgesamt.

## **2 Linguistische Lehre**

Diese Sicht hat Auswirkungen auf die Lehre der Linguistik an den Universitäten. Wenn ich diese im folgenden betrachte und ihnen meine eigenen Überlegungen entgegensetze, so gehe ich davon aus, daß es der Linguistik prinzipiell darum geht, Sprache als kulturelles Objekt zu verstehen, zu beschreiben und zu interpretieren.

### ***2.1 Theorie-gestützte Linguistik***

In den Seminaren zur Linguistik allgemein und zu der der romanischen Sprachen im besonderen wird Sprache allerdings zumeist allein unter theoretischen Gesichtspunkten betrachtet, d.h. es werden bestehende Theorien, von Forschenden durchgeführte Untersuchungen und Beschreibungen in Form von Grammatiken und Wörterbüchern diskutiert. Da die meisten linguistischen Theorien, wie bemerkt, auf einer geringen Zahl isolierter, zumeist erfundener Sätze gründen, die meisten Untersuchungen sich auf funktionelle Aspekte des abstrakten Sprachsystems oder deren Realisierung in der ebenfalls abstrakten Norm konzentrieren und Grammatiken und Wörterbücher aus dem Kontext gerissene Abstraktionen darstellen, spielt die einzige reale Manifestation der Sprache, nämlich das Sprechen und mit Sprechen meine ich sowohl das gesprochene als auch das geschriebene Sprechen in solchen Seminaren eine verschwindend kleine Rolle.

### ***2.2 Daten-gestützte Linguistik***

Nach Wilhelm von Humboldt kann aber Sprache als solche nur in der verbundenen Rede untersucht werden, Grammatiken und Wörterbücher kommen dagegen höchstens ihrem toten Gerippe gleich. Selbst eine vergleichende Studie von nicht sehr umfangreichen und nicht gut gewählten Proben natürlichen Sprechens ermöglicht nach Humboldt mehr Einblick in das Wesen der Sprache als das Studium von Grammatiken und Wörterbüchern:

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

Die Sprache liegt nur in der verbundenen Rede, Grammatik und Wörterbuch sind kaum ihrem toten Gerippe vergleichbar. Die bloße Vergleichung selbst dürftiger und nicht durchaus zweckmäßig gewählter Sprachproben lehrt daher viel besser den Totaleindruck des Charakters einer Sprache auffassen, als das gewöhnliche Studium der grammatischen Hilfsmittel. [...] Freilich führt dies in eine mühevollen, oft ins Kleinliche gehende Elementaruntersuchung, es sind aber lauter in sich kleinliche Einzelheiten, auf welchen der Totaleindruck der Sprachen beruht, und nichts ist mit dem Studium derselben so unverträglich, als bloss in ihnen das Grosse, Geistige, Vorherrschende aufsuchen zu wollen. (cf. Humboldt 1827-1829/1963: 186).

Charles J. Fillmore kommt übrigens zu einem ähnlichen Schluß: "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). Was Humboldt und Fillmore hier vertreten, kann mit einem Terminus von John Sinclair und seiner Schule (see f. ex. Sinclair 1992) als *data-driven linguistics*, als Daten-gestützte Linguistik bezeichnet werden.

### **2.3 Theorie- und datengestützte Linguistik**

Natürlich gibt es auch Seminare, die Theorie und Daten integrieren und wo Studierende ihre Arbeiten auf empirische Untersuchungen stützen. Das Hauptziel ist dabei aber zumeist nicht, vorgefundene Theorien und Modelle auf der Grundlage von systematisch erhobenen Daten kritisch zu hinterfragen, sondern Daten im Rahmen einer bestimmten Theorie zu analysieren und in *a priori* konstruierte Modelle zu pressen. Gelernt werden soll demnach nicht etwas über das natürliche Sprechen, sondern über die Theorie.

Aus dieser Perspektive ist es nur logisch, wenn erwartet wird, daß Untersuchungen auf der Grundlage von Grammatiken, Wörterbüchern, Beispielsammlungen oder relativ kleinen Auszügen von gedruckten oder transkribierten Texten durchgeführt werden und die Datenerhebung lesend und manuell erfolgt. Dieselbe Perspektive ist auch der Grund, warum Computer-gestützte Untersuchungsmethoden bisher, wie eine von mir im Rahmen von ACO\*HUM durchgeführte Studie zu den Internetseiten einer Vielzahl geisteswissenschaftlicher Institute an deutschen Universitäten zeigt, in aller Regel in den Seminaren nicht diskutiert und noch weniger unterrichtet werden und kein Vergeuden von Energie darin gesehen wird, wenn Studierende im Rahmen von Abschlußarbeiten zwar die von ihnen selbst durchgeführten Sprachaufnahmen unter Zurhilfenahme eines PCs transkribieren, dann aber ihre Daten anhand einer ausgedruckten Version der Transkription manuell erheben und analysieren.

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

### **3 Computer-gestützte linguistische Lehre**

All dies befähigt die Studierenden aber weder, Sprache als kulturelles Objekt zu verstehen, das sich allein im Sprechen manifestiert, noch kritisch mit bestehenden Theorien und Ergebnissen von Untersuchungen umzugehen. Dazu bräuchten sie nämlich eine systematische Kenntnis von den wirklichen Daten.

In meinen Augen liegt jedoch der generelle Sinn eines Studiums gerade im Erwerb von Fähigkeiten, die einen fundierten und kritischen Umgang mit existierenden Theorien und Modellen erlauben. Auch linguistische Theorien und Modelle sind ja nicht neutral, sondern Ausdruck einer bestimmten Konzeption von Sprache oder von denjenigen, die sie sprechen. Zudem basieren diese, wie eingangs bemerkt, bisher zumeist auf einer minimalen Datenbasis.

Davon abgesehen baut mein Versuch, synchronische Romanische Linguistik Computer-gestützt zu unterrichten, auf der Überlegung auf, daß ein Studium der romanischen Sprachen immer schon nicht nur spezifische Kenntnisse von der jeweiligen Sprache, sondern auch Schlüsselqualifikationen vermittelt hat, die über das Fachgebiet hinaus von Bedeutung sind. Solche Schlüsselqualifikationen sind: Kenntnisse in der Beschaffung, Verarbeitung, Analyse und Darstellung von Information.

Da es nun neben den traditionellen Informationsquellen und Methoden der Informationsbeschaffung, -verarbeitung, analyse und -präsentation ein neues Medium gibt, wo Information gespeichert und präsentiert wird, einen neuen Typ von Information und neue Werkzeuge zu ihrer Verarbeitung und Analyse, müssen diese Komponenten unserer Realität im Linguistikstudium neben die bisher genutzten treten und zwar nicht nur, weil, wie Henry Rogers ausführt: "Good teachers have a responsibility to use the most advantageous tools available" (Rogers 1998: 63), sondern auch deshalb, weil die Studierenden, die wir heute unterrichten, ihre spezifischen Kenntnisse von den kulturellen Objekten romanische Sprachen und die von ihnen erworbenen Schlüsselqualifikationen nur werden einsetzen können, wenn sie einen Job finden und weil sie einen Job, der ihnen entspricht und sie befriedigt, nur dann finden können, wenn sie, neben allem anderen, kritisch, konstruktiv und selbstbewußt mit den Informationstechnologien umgehen können.

Ich werde nun zunächst das Korpus der romanischen Zeitungssprachen, das dieser Lehre zugrundeliegt, vorstellen. Da die Untersuchungsmöglichkeiten, die ein solches Korpus bietet, maßgeblich vom Markup abhängen, mit dem es angereichert wurde, werde ich auch darauf kurz

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

eingehen. Anhand einer Lehrveranstaltung und einigen Ergebnissen, die Studierende bei dem Versuch erzielten, mit Hilfe von Korpusuntersuchungen Theorie und Empirie zu integrieren, werde ich dann zum Schluß einen Einblick in die Lehrmethode und ihre Auswirkung auf das Lernverhalten der Studierenden und den Ablauf der Lehrveranstaltung zu geben suchen.

### **3.1 Das Korpus romanischer Zeitungssprachen**

Das Korpus der romanischen Zeitungssprache wurde in zwei Phasen für meine eigenen Forschungen erstellt.<sup>1</sup> Da seine Erstellung auf einer bestimmten Theorie von der Sprache und vom Wissen der Sprechenden basiert, die es von anderen Korpora unterscheidet, muß ich diese zunächst charakterisieren. Das Korpus soll, das sei vorweggenommen, keinen *common core* repräsentieren, sondern sich an der Komplexität des natürlichen Sprechens und des Wissens der Sprechenden orientieren.

#### *3.1.1 Das Wissen der Sprechenden in einer heterogenen Sprachgemeinschaft*

Die Theorie, um die es hier geht wurde von Dell Hymes (1972) und Eugenio Coseriu (1988) entwickelt. Beide gehen gerade nicht davon aus, daß Sprachen homogene Einheiten sind, sondern betrachten sie als aus vielen verschiedenen Varietäten zusammengesetzte Entitäten. Unter Sprechen verstehen sie eine sehr komplexe, auf Wissen gründende Aktivität. Die Sprechenden, von denen sie ausgehen, leben in einer realen und deshalb heterogenen Sprachgemeinschaft und besitzen ein sozio-kulturell determiniertes Wissen von mehreren Varietäten und Sprachen. Dell Hymes beschreibt die Situation so: "[...] in much of our world, the ideally fluent speaker-listener is multilingual" und weiter "Even an ideally fluent monolingual of course is master of functional varieties within the one language." (Hymes 1972: 274). Da diese nicht auf eine gemeinsame Grammatik zurückgeführt werden können, wird das Konzept eines *common core* abgelehnt.

Betrachten wir zum Beispiel die Situation in Italien oder Spanien, so existieren hier Dialekte, die von anderen Sprechenden des Italienischen oder Spanischen nicht einfach verstanden werden. Zudem ist der Gebrauch eines bestimmten Dialekts nicht auf seine Ursprungsregion begrenzt, sondern er funktioniert aufgrund von Migration in anderen Regionen oft als sozio-kulturelle oder situative Varietät und kann schon deshalb nicht als regional peripher betrachtet werden. Der Gebrauch von Dialekten ist auch nicht auf niedere Schichten oder Ungebildete

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

beschränkt, er kann also auch nicht als Sprache sogenannter peripherer Klassen vernachlässigt werden. Das heißt, wenn wir dieser Realität bei der Erstellung von Korpora Rechnung tragen wollen, dann müssen wir, Dell Hymes folgend, mit der Tradition brechen, die eine Sprache mit einer Kultur gleichsetzt und von ganz bestimmten Funktionen ausgeht (cf. Hymes 1972: 289).

Darüber hinaus müssen wir auch beachten, daß die Sprechenden nicht nur die von ihnen selbst gesprochenen Sprachen oder Varietäten kennen, sondern auch eine zumindest rudimentäre richtige oder falsche Kenntnis von anderen Sprachen und Varietäten oder zumindest von den Traditionen ihres Imitierens haben.<sup>2</sup> Auch wenn solche Imitationen Elemente enthalten, die von denjenigen, die das Konzept eines *common core* vertreten, nicht in Korpora aufgenommen werden würden, so sind sie doch ein integraler Bestandteil der Sprache der imitierenden Gemeinschaft und nicht der imitierten Sprache. Aus einem Korpus, das sich am realen Wissen der Sprechenden orientieren soll, können also auch die sogenannten Nachahmungssprachen nicht ausgeschlossen werden.

Nicht vergessen dürfen wir zudem, daß das sprachliche Wissen der Sprechenden nicht der Synchronie der Sprache oder der Varietät entspricht, sondern zu jeder Zeit auch Elemente enthält, die wie *thou* im Englischen oder *gebet* und *nehmet* im Deutschen etwa im gegenwärtigen Sprachzustand nicht funktionieren, sondern funktionelle Einheiten eines früheren Sprachzustands repräsentieren, d.h. zur Diachronie der Sprache gehören. Andere Fakten können von den Sprechenden auch in einer diachronen Perspektive bewertet werden, obwohl sie tatsächlich in der aktuellen Synchronie funktionieren. So mögen zum Beispiel einige Sprechende des Italienischen sagen, daß *udire* heute nicht mehr gebraucht wird und durch *sentire* ersetzt worden ist, andere mögen dagegen in *sentire* immer noch eine Innovation sehen.<sup>3</sup>

---

<sup>1</sup> Vgl. hierzu Burr (1993 a u. b, 1994, 1996, 1997).

<sup>2</sup> Italienische Gemeinschaften zum Beispiel, die nicht Toskanisch sprechen, imitieren diesen Dialekt mit Hilfe eines übertriebenen Gebrauchs der sogenannten *gorgia*. In Spanien kann der *seseo* und *yeísmo* ähnlich eingesetzt werden und um das Deutsch nachzuahmen wird an jedes italienische oder englische Wort ein *-en* angehängt. So entstehen dann Formen wie *spaghetten*, *mangiaren* etc. Nachahmungen von Sprachen kommen auch in der Literatur vor. Vgl. hierzu etwa *All's Well That Ends Well*, Akt IV Szenen I & II von Shakespeare. Diese Nachahmungen müssen nicht unbedingt den normalen traditionellen Realisierungen einer bestimmte Varietät oder einer bestimmten historischen Sprache entsprechen. Stattdessen enthalten sie oft sogar Elemente, die in der in Frage stehenden Varietät oder Sprache gar nicht existieren oder sogar unmöglich wären (cf. Coseriu 1988: 148-152). Im Deutschen gibt es zum Beispiel Ausdrücke wie *uno momento*, *picco bello*, *dalli dalli*, *alles paletti*, die Italienisch sein sollen, aber gerade kein Italienisch sind, da sie entweder im Italienischen so nicht existieren oder sogar gegen die Regeln des Italienischen verstoßen.

<sup>3</sup> Ein solches diachronisches Bewußtsein gibt es nach Coseriu vor allem in Sprachgemeinschaften, die eine schriftlich fixierte literarische Tradition besitzen. In solchen Gemeinschaften können wir veraltetete Formen sogar bewußt gebraucht, um uns auf einen älteren Sprachzustand oder einen bestimmten historischen Moment zu beziehen. Formen und die für sie geltenden Normen bleiben also auch dann noch präsent, wenn

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

Da aber für die Sprache normalerweise die Sprechenden, ihre Haltung zu sprachlichen Phänomenen und die Anwendung ihres Wissens ausschlaggebend sind und nicht Linguisten und Linguistinnen (cf. Coseriu 1988: 134-135), kann ein Ausschluß solcher Elemente aus einem Korpus, das sich an den Sprechenden orientiert, nicht gerechtfertigt werden.

### 3.1.2 *Das Korpus*

Da das Korpus romanischer Zeitungssprachen also mit dem Ziel erstellt wurde, von den Sprechenden auszugehen, ist dieses Korpus ausdrücklich heterogener Natur. Das heißt, die Ausgaben der verschiedenen Zeitungen sind so in das Korpus eingegangen, wie sie sich dem Publikum am Tag ihres Erscheinens präsentierten. Eine Abweichung davon stellen lediglich die im Korpus fehlenden Anzeigen dar. Die Integrität aller in den gedruckten Zeitungsexemplaren enthaltenen Texte wurde gewahrt. Auch die Phänomene, die im Falle eines *common core* Korpus das Bild stören würden, wie etwa die in allen Zeitungen enthaltenen dialektalen Elemente oder das Katalanische in der spanischen Komponente oder Elemente, die zur Diachronie der Sprachen gehören, wurden nicht aussortiert. Die einzelsprachlichen Subkorpora stellen also keine homogene Sprache dar, sondern sind ein Porträt der tatsächlichen Kombination von stilistischen und sozio-kulturellen Varietäten in den einzelnen Zeitungen und bilden damit das komplexe sprachliche Wissen ab, das vom Publikum der jeweiligen Zeitung erwartet wird.

Dieses Wissen insgesamt ist allerdings selbst noch eine Abstraktion, denn nicht alle lesen die ganze Zeitung und nicht alle lesen die gleichen Teile. Das heißt, wir müßten noch, um dem Wissen der verschiedenen Schichten näher zu kommen, zwischen ihnen und den einzelnen Teilen der Zeitung anhand der existierenden demographischen Daten eine Relation herstellen. Eine solche herzustellen ist aufgrund des Markup, mit dem das Korpus angereichert wurde, jederzeit möglich.

---

sie nicht mehr zur aktuellen Synchronie der Sprache gehören. So gehört in England etwa die Sprache der St. James's Bible, der Stücke von Shakespeare oder der Romane von Dickens zumindest teilweise zum aktuellen sprachlichen Wissen der Sprechenden. Veraltete Formen können sogar wiederbelebt werden und gehören dann auch wieder zur aktuellen Synchronie (cf. Coseriu 1988: 135-137). Das von Coseriu angeführte Beispiel des spanischen Suffix *-i* kann zur Verdeutlichung hier genügen. Dieses Suffix wurde einmal zur Bildung von Adjektiven für mit dem Orient in Beziehung stehenden Personen oder Dingen verwendet, es wurde dann aber für lange Zeit unproduktiv. Im 20. Jahrhundert wurde es für die Bildung von Adjektiven wie *pakistani* etc. wiederbelebt (cf. Coseriu 1988: 137).

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

### 3.1.3 Die Zusammensetzung des Korpus

Das Korpus setzt sich insgesamt zusammen aus zwei Komponenten: einem Korpus italienischer Zeitungen, die in der Zeit um die «Deutsche Einigung» 1989 veröffentlicht wurden, und einem Korpus französischer, italienischer und spanischer Zeitungen, die zur Zeit der Europawahlen 1994 erschienen sind. Die folgenden Tabellen geben die Zusammensetzung und den Umfang der beiden Komponenten und ihrer Subkomponenten wieder (fig. 1 u. fig. 2):

#### **Korpus «Italienische Zeitungssprache 1989»**

| <b>Zeitung</b>             | <b>Wortformen</b> |
|----------------------------|-------------------|
| <u>Corriere della Sera</u> |                   |
| <b>Gesamt:</b>             | <b>258.287</b>    |
| 19.10.1989                 | 170.129           |
| 20.10.1989                 | 79.030            |
| 21.10.1989                 | 91.099            |
| <u>Il Mattino</u>          |                   |
| <b>Gesamt:</b>             | <b>171.501</b>    |
| 20.10.1989                 | 82.102            |
| 21.10.1989                 | 89.399            |
| <u>La Repubblica</u>       |                   |
| <b>Gesamt:</b>             | <b>174.958</b>    |
| 20.10.1989                 | 88.961            |
| 21.10.1989                 | 85.997            |
| <u>La Stampa</u>           |                   |
| <b>Gesamt:</b>             | <b>119.771</b>    |
| 20.10.1989                 | 69.964            |
| 21.10.1989                 | 49.807            |
| <b>Größe des Korpus</b>    | <b>724.517</b>    |

fig. 1 Korpus 1989: Größe der Subkorpora und ihrer Komponenten

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

### «Korpus romanischer Zeitungssprachen 1994»

| <b>Zeitung</b>                    | <b>Wortformen</b> |
|-----------------------------------|-------------------|
| <u><i>Le Monde</i></u>            |                   |
| <b>Gesamt:</b>                    | <b>236.236</b>    |
| 12./13.06.1994:                   | 89.786            |
| 14.06.1994:                       | 70.936            |
| 15.06.1994:                       | 75.514            |
| <u><i>Corriere della Sera</i></u> |                   |
| <b>Gesamt:</b>                    | <b>303.641</b>    |
| 13.06.1994:                       | 102.976           |
| 14.06.1994:                       | 102.441           |
| 15.06.1994:                       | 98.224            |
| <u><i>La Vanguardia</i></u>       |                   |
| <b>Gesamt:</b>                    | <b>261.133</b>    |
| 13.06.1994:                       | 86.468            |
| 14.06.1994:                       | 94.251            |
| 15.06.1994:                       | 80.414            |

fig. 2 Korpus 1994: Größe der Subkorpora und ihrer Komponenten

#### 3.1.4 Das Markup-System

Jede Komponente der beiden Korpora wurde mit dem im folgenden aufgeführten Markup-System im COCOA-Format angereichert (fig. 3). Dieses hält zunächst bibliographische Informationen, dann die Struktur der Zeitung und schließlich die intendierte Art des Sprechens innerhalb der Texttypen fest.<sup>4</sup> Aufgrund dieser Kodierung sind eine ganze Reihe von Unterscheidungen hinsichtlich der Schichten des Publikums und der sprachlichen Varietäten möglich.

---

<sup>4</sup> Die Entwicklung des Systems wurde im Rahmen der Erstellung des «Korpus der italienischen Zeitungssprache 1989» begonnen. Dabei wurden die in der italienischen Presse üblichen Namen für die Werte der Variablen übernommen.

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

| <b>Einheit</b>                  | <b>Variable</b> | <b>Beispiel</b>   |
|---------------------------------|-----------------|-------------------|
| Zeitung                         | <Z>             | <Z La Vanguardia> |
| Ausgabe                         | <E>             | <E 130694>        |
| Sparte                          | <S>             | <S Politica>      |
| Ausgewiesenheit des Textes      | <A>             |                   |
| signiert                        |                 | <A firmato>       |
| anonym                          |                 | <A non firmato>   |
| Autor/Autorin                   | <N>             | <N Tapia Juan>    |
| Seite                           | <C>             | <C 01>            |
| Sprache                         | <L>             | <L Inglese>       |
| Textart                         | <T>             |                   |
| Vorzeile                        |                 | <T Occhiello>     |
| Schlagzeile                     |                 | <T Titolo>        |
| Untertitel                      |                 | <T Sottotitolo>   |
| Zwischenüberschrift             |                 | <T Catenaccio>    |
| Zusammenfassung                 |                 | <T Sommario>      |
| Ankündigung                     |                 | <T Civetta>       |
| Artikel                         |                 | <T Articolo>      |
| 'Aufmacher'                     |                 | <T Spalla>        |
| Leitartikel                     |                 | <T Fondo>         |
| Glosse                          |                 | <T Corsivo>       |
| Kolumne                         |                 | <T Rubrica>       |
| Kritik                          |                 | <T Critica>       |
| Interview                       |                 | <T Intervista>    |
| Nachricht                       |                 | <T Notizia>       |
| Kurznachricht                   |                 | <T Breve>         |
| Kurzmeldung                     |                 | <T Flash>         |
| Leserbrief                      |                 | <T Lettera>       |
| Liste                           |                 | <T Elenco>        |
| Fernseh-, Kinoprogramm          |                 | <T Programma>     |
| Filminhalt                      |                 | <T Film>          |
| Wetterbericht                   |                 | <T Tempo>         |
| Bildunterschrift                |                 | <T Foto>          |
| Buch-, Film-, Liedtitel, etc.   |                 | <T Nome>          |
| Art des Sprechens               | <P>             |                   |
| fortlaufender Text              |                 | <P Prosa>         |
| Zitat von schriftlichen Quellen |                 | <P Citazione>     |
| Zitate von mündlichen Quellen   |                 | <P Discorso>      |
| Frage im Interview              |                 | <P Domanda>       |
| Antwort im Interview            |                 | <P Risposta>      |

fig. 3 Das Markup-System

Im älteren italienischen Korpus wurden zudem alle finiten Verbformen mit einem Zahlenkode entsprechend den Kategorien Tempus, Modus und Aspekt, die sie aufgrund einer bestimmten Theorie realisieren, kodiert. Der Kode legt zunächst einmal fest, daß analytische Formen (stavo#I210a011cantando) wie synthetische (I130a003canteranno) Formen jeweils

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

als Einheit betrachtet werden (#). \$I210a011 gibt an, daß es sich um eine Verbform im Modus Indikativ (I), die die inaktuelle Ebene (2) und die parallele (10) Perspektive realisiert, die im Aktiv (a) erscheint und eine Verbalperiphrase vom Typ *stare* + *Gerundium* (01) in der ersten Person (1) ist. Bei \$I130a003 handelt es sich dagegen um eine Verbform im Indikativ (I), die die aktuelle Ebene (1) und die prospektive Perspektive (30) realisiert, die im Aktiv erscheint (a), keine Verbalperiphrase ist (00) und in der 3. Person steht (3).<sup>5</sup>

Aufgrund dieses hier vorgestellten Markup ist es möglich, anhand des Korpus nicht nur die Varietäten, aus denen das Korpus sich zusammensetzt, sondern auch grammatische und stilistische Phänomene zu untersuchen. Da die Zeitungen im romanischen Korpus zudem alle zur gleichen Zeit erschienen sind, lassen sich auch synchronische Vergleiche zwischen den drei romanischen Sprachen anstellen. Das italienische Korpus erlaubt, da es aus zwei zu unterschiedlichen Zeiten zusammengestellten Subkorpora besteht, darüber hinaus auch diachronisch ausgerichtete Untersuchungen.

### 3.1.5 Das Korpus unter TACTWeb

Das Projekt eines Computer-gestützten Linguistikunterrichts wurde 1990 begonnen und durchlief bisher eine Reihe von Phasen, in denen unterschiedliche Software zum Einsatz kam (*OCP*, *Logiciel St. Chef*, *DBT* und *Tact 2.1*). Die damit gemachten Erfahrungen führten zu der Erkenntnis, daß eine Computer-gestützte linguistische Lehre einer interaktiven und zugleich Hardware-unabhängigen Analysemöglichkeit bedarf. Eine solche konnte schließlich mit Hilfe von TACTWeb und der Einspeicherung eines Teils des zu Datenbanken kompilierten Korpus in das Internet geschaffen werden. Für eine online-Analyse stehen damit nun unter [www.uni-  
duisburg.de/FB3/ROMANISTIK/PERSONAL/Burr/humcomp/home.htm](http://www.uni-duisburg.de/FB3/ROMANISTIK/PERSONAL/Burr/humcomp/home.htm) die folgenden Teilkorpora zur Verfügung:

| <b>Zeitung</b>                | <b>Ausgabe</b> | <b>Wortformen</b> |
|-------------------------------|----------------|-------------------|
| <i>Il Corriere della Sera</i> | 20.10.1989     | 79.030            |
| <i>Il Mattino</i>             | 20.10.1989     | 82.102            |
| <i>Le Monde</i>               | 15.06.1994     | 75.514            |
| <i>Il Corriere della Sera</i> | 15.06.1994     | 98.224            |
| <i>La Vanguardia</i>          | 15.06.1994     | 80.414            |

<sup>5</sup> Für eine genaue Darstellung des zugrundeliegenden Verbalsystems und der Kodierung vgl. Burr (1993: 75-110,157-161, 467-468).

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

fig. 4 Online-Korpuskomponenten

### **3.2 Ein Beispiel**

Mein Vorgehen bei der Computer-gestützten Lehre der Linguistik will ich zum Schluß noch an einem Beispiel illustrieren. Es handelt sich dabei um ein Seminar zu "Tempus, Aspekt, Modus (Französisch, Italienisch & Spanisch)", das ich im Wintersemester 1998/99 an der Universität Siegen angeboten haben.

#### *3.2.1 Das Vorgehen*

Da nur das «Korpus italienische Zeitungssprache 1989» eine Kodierung der finiten Verbformen aufweist, mußten das französische und spanische Subkorpus von den Studierenden selbst im Sinne der von ihnen durchzuführenden Analyse aufbereitet werden, d.h. sie mußten die Realisierungen bestimmter Verbalkategorien kodieren. Modell stand dabei das oben dargestellte Kodierungssystem. Die kodierten Teilkorpora wurden dann, in Datenbanken transformiert, jeweils unter dem Namen der oder des Studierenden über das Internet von der TACTWeb-Korpusseite aus zugänglich gemacht. Sie sind auch weiterhin dort verfügbar.

Der erste Teil des Seminars war der Einführung in das Thema und der Einarbeitung in die von TACTWeb gebotenen Analysemöglichkeiten gewidmet. Dabei mußten die Studierenden, neben der Vorbereitung des Korpus für die empirische Untersuchung, eine Reihe von Aufgaben erfüllen. Zunächst war eine thematisch relevante Bibliographie zu erstellen und nach einer ersten Lektürephase ein Projektvorschlag einzureichen. Der mündliche Seminarbeitrag bestand aus einem Bericht über die bisherige Forschung zu der jeweiligen Kategorie/zu den jeweiligen Kategorien. Zur Diskussion gestellt wurden dabei auch die vorläufige Struktur der späteren schriftlichen Hausarbeit sowie mögliche Untersuchungsschritte. Da die empirische Untersuchung integrativer Bestandteil der Lehrveranstaltung sein sollte, konzentrierte sich der zweite Teil des Seminars v.a. auf die Durchführung der eigentlichen Analyse. Das Seminar wurde mit einer schriftlichen Hausarbeit abgeschlossen, die neben der bisherigen Forschung auch die Analyseschritte und die bei der empirischen Untersuchung erzielten Ergebnisse diskutieren sollte.

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

### 3.2.2 Einige Ergebnisse

Beim Vergleich des synthetischen mit dem periphrastischen Futur (*aller faire*) im französischen Subkorpus kommt eine Studentin aufgrund ihrer genauen Analyse aller Belegstellen im Korpus zu dem Schluß, daß mit ersterem eine Verbalhandlung im allgemeinen neutral dargeboten, mit letzterem dagegen eine persönliche Beteiligung oder Nähe ausgedrückt wird. Das folgende Beispiel ist hierfür besonders indikativ:

Selon M. Balladur, 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%. (Le Monde 15.06.1994).

Die Untersuchung zum Gebrauch des *Perfecto simple* und *Perfecto compuesto* in den Sparten *Política*, *Internacional*, *Deportes* des spanischen Teilkorpus, die eine Studentin des 2. Studienjahrs durchgeführt hat, liefert als Antwort auf die Frage, ob *perfecto compuesto* in Zitaten mündlicher Quellen häufiger gebraucht wird als im fortlaufenden Text, das folgende Ergebnis:

| Art des Sprechens        | perfecto compuesto | perfecto simple |
|--------------------------|--------------------|-----------------|
| fortlaufender Text:      | 23%                | 77%             |
| Zitat mündlicher Quelle: | 67%                | 33%             |

Ein weiterführender Untersuchungsschritt zu den Unterschieden zwischen den drei Sparten ergibt:

| Sparte         | perfecto compuesto | perfecto simple |
|----------------|--------------------|-----------------|
| Deportes:      | 18                 | 20              |
| Política:      | 27                 | 3               |
| Internacional: | kaum Zitate        |                 |

Die Studentin interpretiert dieses Ergebnis folgendermaßen. In der Sparte *Política* geht es am 15.06.1994 vor allem um die Europawahlen. Dabei kommen viele Politiker und Politikerinnen zu Wort, die die Wahlergebnisse und die möglichen Folgen kommentieren. Der Tempusgebrauch steht damit in Zusammenhang, d.h. er drückt, wie die folgenden beiden Beispiele zeigen, die temporale und emotionale Nähe der Ereignisse aus:

Los ciudadanos han hablado.

Se ha abierto una etapa nueva (La Vanguardia 15.06.1994).

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

Eine weitere Untersuchung fragt nach dem Zusammenhang zwischen dem Gebrauch von *Passato remoto* und der diatopischen Variation.<sup>6</sup> Sie kommt u.a. zu dem Ergebnis, daß sich der diatopische Unterschied nicht am Anteil von *passato remoto* Formen an der Gesamtzahl finiter Verbformen (*Mattino* 2% - *Corriere* 3,2%) festmachen läßt, sondern sich v.a. im Bereich der Kopula zeigt. Sie erscheint nämlich im *Mattino* zumeist im *passato remoto* (*fu/fummo/furono*) im *Corriere* dagegen im Imperfekt (*era/eravamo/erano*):

[...] per non dire di quel Morse che fu la bandiera dei successi Ignis. (*Il Mattino* 20.10.1989)

Morse, in fondo, era soprattutto un gran tiratore. (*Il Corriere della Sera* 20.10.1989)

#### 4 Lernen und Lehren

Wie der obige Einblick zeigt, lassen sich Computer-gestützte Lehre und traditionelle Seminare positiv integrieren.<sup>7</sup> Ein unter TACTWeb verfügbares Korpus weckt zudem, da es erlaubt, zunächst einmal mit Abfragen und den dabei erhobenen Daten zu spielen, das Interesse der Studierenden am Funktionieren von Sprache. Dabei lernen sie schnell, ihre Suchen genauer zu definieren, Hypothesen aufzustellen und nach Beweisen zu suchen.

Gleichzeitig lernen sie, was Korpora eigentlich sind, was das Anreichern mit Markup bedeutet und welche Untersuchungen dadurch überhaupt erst möglich werden bzw. welches die Grenzen sind. Dadurch, daß Untersuchungen an einem solchen Korpus jederzeit relativ schnell wiederholt bzw. anders aufgebaut werden können, werden sie animiert, weiterführende Fragen auch dann noch zu stellen, wenn ihre Untersuchung schon weit fortgeschritten ist, können also auch erst später Bedachtes noch für ihre Untersuchung nutzen. Zudem erlauben ihnen die systematisch erhobenen Daten und die darin zu Tage tretenden Regelmäßigkeiten, zumindest ansatzweise, Theorien und von anderen durchgeführte Untersuchungen kritisch zu hinterfragen.

Auch auf die Seminarsituation wirkt sich dieses Vorgehen positiv aus. Aufgrund der in den Sitzungen selbst unternommenen Untersuchungen kommt es nämlich generell zu viel mehr Kommunikation unter Studierenden und Lehrenden. Auch wächst das Interesse an den

---

<sup>6</sup> Im Gemeinitalienischen gibt es nach einhelliger Meinung drei regional bedingte Tempussysteme, die sich vor allem anhand des *passato remoto* unterscheiden. Um die Untersuchung diatopischer Unterschiede zu erlauben, gingen in das «Korpus italienische Zeitungssprache 1989» Zeitungen aus unterschiedlichen Gegenden Italiens ein. Denkbar wäre, daß in *Il Mattino* (Neapel) mehr *passato remoto* Formen erscheinen als in *Il Corriere della Sera* (Mailand).

<sup>7</sup> Eine solche Integration wird von Aimée Morrison (1999) in ihrem sehr interessanten Beitrag auch für die literaturwissenschaftliche Lehre gefordert.

Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): Sprache und Kultur (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.

Referaten, am Stellen von Fragen und an der Diskussion. Das wiederum fördert den Wissenstransfer. Die Lehrende hat zudem mehr Zeit zur Betreuung einzelner Studierenden, erhält eher Einblick in Lesegänge, Probleme und in den Fortschritt des empirischen Projekts.

Das Allerwichtigste ist aber, wie mir scheint, daß die Studierenden zum einen erfahren, daß Sprechen uns als verbundene Rede entgegentritt und nicht in Einzelementen, zum anderen Schlüsselqualifikationen erwerben, die sie in vielen Betätigungsfeldern einsetzen können.

## **BIBLIOGRAPHIE**

- Burr, Elisabeth (1993a): *Corpus of Italian newspapers*. Oxford: Oxford Text Archive U\*-1723-D.
- Burr, Elisabeth (1993b): *Verb und Varietät*. Ein Beitrag zur Bestimmung der sprachlichen Variation am Beispiel der italienischen Zeitungssprache (= Romanistische Texte und Studien 5). Hildesheim/Zürich/New York: Olms.
- Burr, Elisabeth (1994): „Italian Newspaper Corpus (ita03)“, in: ECI/MCI: *European Corpus Initiative Multilingual Corpus 1*. CD-ROM: \data\eci1\.
- Burr, Elisabeth (1996): „A Computer Corpus of Italian Newspaper Language“, in: *Research in Humanities Computing 4*. Selected Papers from the ALLC/ACH Conference, Christ Church, Oxford, April 1992. Oxford: Clarendon Press 216-239.
- 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 (1988): *Sprachkompetenz: Grundzüge der Theorie des Sprechens* (= UTB 1481). Tübingen: Francke.
- 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.
- Gross, Maurice (1994): „Constructing Lexicon-Grammars“, in: Atkins, Sue B.T./Zampolli, Antonio (eds.): *Computational Approaches to the Lexicon*. Oxford: Oxford University Press 213-263.
- 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.

- Burr, Elisabeth (2000): "Das Korpus romanischer Zeitungssprachen und sein Online-Einsatz in der Lehre", in: Schlosser, Horst Dieter (ed.): *Sprache und Kultur* (= forum Angewandte Linguistik 38). Frankfurt am Main / Berlin / Bern / Bruxelles / New York / Oxford / Wien: Lang 137-151.
- Hymes, Dell H. (1972): „On Communicative Competence“, in: Pride, J. B. / Holmes, Janet (eds.): *Sociolinguistics* (= Penguin Modern Linguistics Readings). Harmondsworth: Penguin 269-293.
- Morrison, Aimée (1999): "Teaching Humanities Computing ... A Graduate Student Perspective", <http://www.humanities.ualberta.ca/amorrison/otherwork/cosh-coch99.htm>.
- Rogers, Henry (1998): "Education", in: Lawler, John / Dry, Helen Aristar (eds.): *Using Computers in Linguistics. A Practical Guide*. London: Routledge 63-100.
- Sinclair, John (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- 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.
- Velde, Roger G. van de (1979): „Probleme der linguistischen Theoriebildung einer empirischen Textwissenschaft“, in: Bergenholtz, Henning/Schaeder, Burkhard (eds.): *Empirische Textwissenschaft. Aufbau und Auswertung von Text-Korpora* (= Monographien Linguistik und Kommunikationswissenschaft 39). Königstein/Ts.: Scriptor 10-27.