

Burr, Elisabeth (1998): "Elektronisches Korpus romanischer Zeitungssprachen", in: *Unsere Duisburger Hochschule* 50, 2: 74-75.

Elektronisches Korpus romanischer Zeitungssprachen

Elisabeth Burr

Über die letzten zehn Jahre wurde von Frau Burr eine große Menge an französischen, italienischen und spanischen Zeitungsausgaben in elektronischer Form zusammengetragen, indem sie entweder, wie in den Anfangsjahren, als sie im Rahmen ihres Promotionsprojektes ein Korpus zur italienischen Zeitungssprache erstellt hat, den Text ganzer Zeitungsexemplare eingescannt, oder wie bei dem im Rahmen ihres Habilitationsprojektes entstandenen Korpus der französischen, italienischen und spanischen Zeitungssprache die Zeitungsausgaben in elektronischer Form direkt von den Verlegern oder als CD-ROM-Ausgaben erworben hat. Thematisch und zeitlich ist das erste Korpus um die Wiedervereinigung 1989, das zweite Korpus um die Europawahlen 1994 organisiert, so daß in beiden Korpora Berichterstattungen zu einem international gleichmäßig wichtigen und interessanten Ereignis zu finden sind, die nicht nur sprachlich, sondern auch inhaltlich eine eigene Untersuchungseinheit bilden.

Bisher konnte zwar nur ein relativ kleiner Teil des gesamten Materials so aufbereitet und mit Tags angereichert werden (vgl. fig. 1), daß es internationalen Standards entsprechend als Korpus eingeordnet werden kann, schon jetzt wird aber das bisher der wissenschaftlichen Öffentlichkeit über das Oxford Text Archive zur Verfügung gestellte Teilkorpus der italienischen Zeitungssprache von 1989, das ebenso wie das Korpus von 1994 eine dreiviertel Million Wörter umfaßt, in vielen Ländern der Welt für Unterrichts- und Forschungszwecke eingesetzt.

Welche Möglichkeiten diese Korpora für die Untersuchung der heutigen romanischen Sprachen bergen, haben Frau Burrs eigene Forschungen sowie die auf dem italienischen Teil des Korpus von 1994 basierende Magisterarbeit von Frau Bettina Möllers gezeigt. Diese Möglichkeiten sind ganz maßgeblich von der die Komplexität der Zeitungssprache respektierenden und systematischen Art der Aufarbeitung des Materials bedingt, die national und international sehr große Anerkennung findet.

Burr, Elisabeth (1998): Elektronisches Korpus romanischer Zeitungssprachen", in: *Unsere Duisburger Hochschule* 50, 2: 74-75.

Bezug	Kodierung	Beispiel
Zeitung	<Z>	<Z La Vanguardia>
Ausgabe	<E>	<E 130694>
Sparte	<S>	<S Politica>
Ausgewiesenheit des Textes	<A>	
signiert		<A firmato>
anonym		<A non firmato>
Name der Autorin/des Autors	<N>	<N Tapia Juan>
Seite	<C>	<C 01>
Sprache	<L>	<L Inglese>
Textart	<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>
Fernseh-, Kinoprogramm		<T Programma>
Filminhalt		<T Film>
Glosse		<T Corsivo>
Interview		<T Intervista>
Kolumne		<T Rubrica>
Kritik		<T Critica>
Kurzmeldung		<T Flash>
Kurznachricht		<T Breve>
Leitartikel		<T Fondo>
Leserbrief		<T Lettera>
Liste		<T Elenco>
Nachricht		<T Notizia>
Wetterbericht		<T Tempo>
Buch-, Film-, Liedtitel, etc.		<T Nome>
Bildunterschrift		<T Foto>
Art des Sprechens	<P>	
fortlaufender Text		<P Prosa>
Zitat von schriftlichen Quellen		<P Citazione>
mündliches Sprechen		<P Discorso>
Frage im Interview		<P Domanda>
Antwort im Interview		<P Risposta>

fig. 1 Das Markup-System

Nicht nur letztere, sondern auch der jetzt, wo die Sprachwissenschaften immer mehr den Wert von systematischen Erhebungen von Sprachdaten erkennen, besonders spürbare Mangel an frei verfügbaren, d.h. nicht kommerziell eingebundenen elektronischen Korpora der romanischen Sprachen hat Frau Burr in ihrer Absicht bestärkt, das gesamte vorliegende Material im Rahmen eines Drittmittelprojektes zu einem großen Korpus der romanischen Zeitungssprachen zu verarbeiten und es zusammen mit einem ausgefeilten und für sprachwissenschaftliche Untersuchungen geeignetem Analyseprogramm auf einer CD der wissenschaftlichen Öffentlichkeit zugänglich zu machen. Modell steht hierfür u.a. auch das British National Corpus (BNC).

Burr, Elisabeth (1998): Elektronisches Korpus romanischer Zeitungssprachen", in: *Unsere Duisburger Hochschule* 50, 2: 74-75.

Daß an einem solchen Antrag auf ein Drittmittelprojekt derzeit aber auch tatsächlich intensiv gearbeitet werden kann, ist nur der von der Gerhard-Mercator-Universität GH bewilligten Anschubförderung zu verdanken. Erst durch diese Förderung wurde es nämlich möglich, mit Frau Möllers nicht nur eine qualifizierte, sondern, da sie im Rahmen ihrer Magisterarbeit schon eine Untersuchung anhand eines Teils des Korpus durchgeführt hat und jetzt ein Korpus-basiertes Promotionsprojekt plant, auch eine überaus interessierte und geeignete wissenschaftliche Hilfskraft einzustellen. Mit ihr zusammen konnte dann u.a. auch die internationale Lage im Bereich von Sprachkorpora auf dem von der *European Language Resources Association (ELRA)* im Mai diesen Jahres in Granada abgehaltenen internationalen Kongreß "First International Conference on Language Resources and Evaluation" zu sondiert werden. Bei diesem Anlaß konnten zugleich bestehende Kontakte zu anderen Korpusprojekten und der *Text Encoding Initiative (TEI)*, die einen SGML-Standard für die Kodierung von Korpora entwickelt hat, vertieft und die Weichen für eine Zusammenarbeit auf europäischer Ebene gestellt werden.