

Diskussion „Transkriptsammlung“

Folgendes ist ein Ausschnitt einer Diskussion, die zwischen dem 17.02. und 21.02.2003 auf der Mailingliste „Gesprächsforschung“ geführt wurde.

Die Mailliste zur Gesprächsforschung

Beiträge senden an: mailliste@gespraechsforschung.de
Abmelden / Adresse ändern unter <http://www.gespraechsforschung.de/liste.htm>

Liebe KollegInnen,

gerade raufe ich mir alle Haare aus beim Lesen einer Diss, die ich "zweitbegutachten" muss. Da hat jemand zwei Jahre lang Videoaufnahmen von Gesprächen zwischen insgesamt 3 Müttern und ihren Kindern gemacht, die brav transkribiert und analysiert und wollte herausfinden, welchen Einfluss das mütterliche Kommunikationsverhalten auf die kindliche Sprachentwicklung hat. Nun war es leider so, dass das eine Kind eine berufstätige Mutter hatte, die kaum Zeit hatte, und sonst niemanden, ein Kind hatte Geschwister, eins war der absolute Mittelpunkt einer umfangreichen erwachsenen Verwandtschaft und hatte ständig Zuspruch. Oh Wunder, dieses Kind hatte die beste Sprachentwicklung, und ich würde nun nicht unbedingt sagen, dass das Kommunikationsverhalten der Mutter die alleinige Ursache war. Ich weiß, dass ich das auch meinem Frisör erzählen könnte, aber können wir nicht etwas dagegen unternehmen, dass ständig derartige sinnlose Fleißarbeiten gemacht werden, die nichts aussagen können? Vielleicht findet sich ja unter Ihnen/Euch jemand, der mal die zahllosen so entstandenen Korpora sammelt und für weitere Arbeiten zugänglich macht, damit nicht jeder wieder bei Null anfängt? Oder vielleicht gibt es so etwas sogar schon in Ansätzen, nur ist es zu wenig bekannt?

Es wäre toll, wenn wir da unsere Forschung etwas effektivieren könnten.

Lieber Herr Kollege,

[...]
es ist nun einmal so, dass wenn jede(r) wieder anfängt, Daten zu erheben, in Anbetracht der zur Verfügung stehenden Zeit nie wirklich aussagefähige Analysen entstehen können. Das führt dann dazu, dass die Arbeiten durchaus gute Noten bekommen, wenn sie nach den Regeln der Kunst gemacht sind, aber die darein gesteckte Arbeit führt nur zu 2 Buchstaben mehr am Namen des Verfassers/der Verfasserin, die Forschung kommt kein Stück weiter. Das ist das, was mich nervt und was durchaus sehr oft auch die betroffenen DoktorandInnen nervt (wir hatten in diesem konkreten Fall im Vorfeld über dieses Problem gesprochen, aber es war kein geeignetes Korpus zum Auswerten zu finden). Wenn sich jemand finden würde, der solche Video-/Transkriptsammlungen von verschiedenen Verfassern wissenschaftlicher Untersuchungen zu solchen Themen zusammenstellt und für Forschungszwecke bereitstellt unter den üblichen Bedingungen zum Schutz der beobachteten Personen, könnten wir erheblich professioneller arbeiten.

Liebe Liste,

Vielleicht kann ich ja Ihre Anfrage...

- > findet sich ja unter Ihnen/Euch jemand, der mal die zahllosen so
- > entstandenen Korpora sammelt und für weitere Arbeiten
- > zugänglich macht, damit nicht jeder wieder bei Null anfängt?
- > Oder vielleicht gibt es so etwas sogar schon in Ansätzen, nur ist es zu
- > wenig bekannt?
- > Es wäre toll, wenn wir da unsere Forschung etwas effektivieren könnten.

... als Anlass für eine Frage nehmen, die ich der Liste schon länger mal gerne gestellt hätte: Es gibt "so etwas" in der Tat schon in Ansätzen. Z.B. kümmern sich die Projekte "OLAC" (<http://www.language-archives.org/>) und "Talkbank" (<http://www.talkbank.org/>) um eine technische und organisatorische Infrastruktur zur Archivierung und zum Austausch von sprachlichen Ressourcen. Wirklich Vergleichbares im deutschsprachigen Raum gibt es m.W. nicht, und die Wahrnehmung oder das Interesse an den genannten Initiativen scheint mit von deutscher Seite auch nicht gerade überwältigend zu sein. Ist der Eindruck so richtig? Wenn ja, was ist der Grund für das mangelnde Interesse deutscher GesprächsforscherInnen/LinguistInnen an diesem Thema?

Lieber [...],

ich habe den Mailwechsel (der durch Frau Albert ausgegangen ist) mitverfolgt. Eine mögliche Antwort auf deine Frage, warum das Bedürfnis nach einer Video- und Transkriptsammlung offensichtlich kaum besteht, ist vermutlich der Umstand, dass einem beim "fremden" Material einfach wichtige Informationen fehlen, die der/die Filmende/Aufnehmende durch diese Arbeit automatisch aufnimmt, nicht immer aber ausreichend zusätzlich schriftlich dokumentiert bzw. dokumentieren kann - vieles läuft da unbewusst im Umfeld ab.

Meine Erfahrungen - ich habe einige Hochschulseminar-Einheiten analysiert - waren aber auch, dass eine Weitergabe bzw. Veröffentlichung des Materials von den Beteiligten ausdrücklich abgelehnt wird. Das könnte auch ein möglicher Grund sein.

Die Fallstudien leisten m.E. sehr wohl einen wichtigen, nicht zu umgehenden Beitrag in der Forschung und eine Zusammenführung der Ergebnisse aus bereits abgeschlossenen Analysen ist nicht weniger effektiv.

soweit meine Gedanken

liebe kollegInnen,

es ist wirklich eine interessante frage, warum keine sammlung zugänglicher transkripte zustande kommt und jede/r von null beginnen muss und sicherlich liegt es nicht nur an der organisation und dem arbeitsaufwand, der damit verbunden ist.

als ich vor einiger zeit bei [...] in dänemark war, bat er mich sofort um die TS zum gehaltenen vortrag, denn auf der www.conversation-analysis.net seite können TS heruntergeladen werden. in dänemark scheint dies kein solch ein großes problem zu sein und er bemerkte, dass es ein spezifisches phänomen in deutschland sei, die daten nicht weiterzugeben. und wenn ich meine eigene praxis kritisch betrachte, stelle ich - obwohl ich mich in der liste stets für einen datenpool ausgesprochen habe - in der tat selbst einige bedenken fest, die ich hier einmal öffentlich machen möchte, da ich denke, dass ich damit nicht alleine bin. vielleicht regt das ja zur diskussion und öffnung an.

ich vermute folgende gründe für das nicht-zustandekommen einer solchen sammlung:

1. ich vermute, die meisten haben - wie ich - eine unsicherheit über die "korrektheit" der transkription (in den augen anderer). wir haben am lehrstuhl [...] bei einem transkriptionsvergleich (bei dem wir jeder dieselbe TS-stelle zu hause unabhängig voneinander transkribiert hatten) sage und schreibe so viele transkriptvarianten wie personen zustande gebracht , von den heißen diskussionen danach, wie was zu hören sei, einmal ganz abgesehen..... und das bei einem (erfahrenen) team [...], das zudem seit jahren eng zusammengearbeitet hatte und über jahre wöchentliche datensitzungen zusammen bestritten hatte. alle, die wir mit TS zu tun haben, werden natürlich auf einer abstrakten ebene zugeben, dass es das "korrekte" TS nicht gibt, aber im konkreten fall, scheint eine gesichtsbedrohung damit verbunden zu sein.

2. der nächste grund, den ich vermute, betrifft die vertraulichkeit der datenbehandlung, die ich bisher immer bei aufzeichnung/aufnahme persönlich, verbindlich zusagen musste. da fällt es einem schwer, selbst ein nicht zu identifizierendes TS aus der hand zu geben, weil man die weitere verwendung nicht kontrollieren kann - auch wenn es irrational ist. das entgegengebrachte konkrete vertrauen stellt eine persönliche verbindlichkeit her, die schwer zu überwinden ist. wenn mich kollegInnen, die ich kenne, nach ausschnitten frage, stelle ich diese gerne zu verfügung, das ist dann irgendwie überschaubarer. aber so anonym.....irrational, aber wirkmächtig.

3. der letzte grund, der mir einfällt, scheint mir zugleich jedoch der stärkste zu sein: konkurrenz. viel arbeit und zum teil jahrelange (!!) vorbereitung steckt in der "ergatterung" von daten, von der transkription ganz zu schweigen und der eigene korpus ist ein pfund , mit dem man z.T. (zumindest wenn ich unsere datensätze anschau [...]) lange wuchern muss und will. das stellt sich nicht so leicht zur verfügung, zumindest nicht, wenn man noch dran arbeitet und das kann ja - wie gesagt - jahre sein.

nun, ich hoffe mit dieser offenbarung von inneren widerständen eine diskussion anzuregen und freue mich auf weitere beiträge
mit kollegialen grüßen

Liebe Liste,

Bislang wurden vier Gründe genannt, die das mangelnde Interesse an einer Video-/ und oder Transkriptsammlung zu erklären versuchen:

- 1) Die Transkriptionen anderer sind für die eigene Arbeit nur begrenzt hilfreich, weil nicht alle zu ihrem Verständnis notwendigen Angaben dokumentiert sind.
- 2) Mit der Veröffentlichung von Transkriptionen setzt man sich der Kritik an der Zuverlässigkeit derselben aus.
- 3) Die Veröffentlichung von Transkriptionen/Aufnahmen ist aus datenschutzrechtlichen Gründen problematisch.
- 4) Die Veröffentlichung von Transkriptionen/Aufnahmen kann einen Nachteil im wissenschaftlichen Wettbewerb darstellen.

Alle vier Gründe sind sicherlich zunächst einleuchtend und nachvollziehbar. Ich glaube aber, dass es zu jedem einzelne auch gute Gegenargumente gibt, und - was vielleicht interessanter ist - dass anderswo eine Auseinandersetzung mit ihnen stattfindet, die hierzulande vielleicht wiederum nicht ausreichend bekannt ist:

zu 1): Steven Bird hat einen Aufsatz mit dem Titel "Seven Dimensions of Portability for Language Documentation and Description" (<http://www ldc.upenn.edu/sb/home/publications.html>) geschrieben, der sich zur einen Hälfte mit den technischen Voraussetzungen für den Austausch von Sprachressourcen befasst, zur anderen Hälfte genau dieses Problem einer brauchbaren Dokumentation von Sprachressourcen thematisiert. Verschiedene Initiativen haben darüber hinaus konkrete Vorschläge für eine Menge von Meta-Informationen erarbeitet, mit deren Hilfe sich Transkriptionen in einer für andere brauchbaren Weise charakterisieren lassen. Siehe z.B. die Vorschläge der schon genannten OLAC (<http://www.language-archives.org/OLAC/olacms.html>) oder von EAGLES/ISLE (<http://www.mpi.nl/ISLE/>)

zu 2): Das ist wohl unbestreitbar ein heikles Problem, aber vielleicht eines, dem man nicht durch eine Nicht-Veröffentlichung von Transkriptionen ausweichen sollte. Ich könnte mir sogar ganz im Gegenteil vorstellen, dass der Weg zu verlässlicheren Transkriptionen nur über die Veröffentlichung derselben (und natürlich der zugehörigen Aufnahmen) führt, weil nur so eine vernünftige Diskussion über die konkreten Schwierigkeiten möglich wird.

zu 3): Dieses Argument habe ich selbst am allerhäufigsten gehört. Ich habe den Eindruck, dass ihm zu großen Teilen ein Missverständnis zugrunde liegt: Die Nutzung des Internets zur Distribution von Sprachressourcen muss nicht gleichbedeutend mit einem unkontrollierten Zugriff durch Jedermann sein. Ein Zugang zu Transkriptionen und Aufnahmen kann passwortgeschützt erfolgen, verantwortlich für die Zuweisung eines solchen Passworts kann z.B. der Urheber der Aufnahme oder die aufgenommene(n) Person(en) sein. Um eine noch größere Sicherheit zu haben, müssen gar die Transkriptionen selbst gar nicht auf einem Server abgelegt werden - es wäre schon ein Fortschritt, wenn sich deren Existenz über das Internet recherchieren ließe. Brian MacWhinney hat für das Talkbank-Projekt einen Vorschlag ausgearbeitet, nach dem Aufnahmen und Transkripte einem von neun "Access Levels" zugeordnet werden (<http://www.talkbank.org/share/levels.html>). Nur auf der untersten dieser Stufen wäre ein freier, unkontrollierter Zugang möglich, auf der "höchsten Sicherheitsstufe" wäre eine Einsicht in die Daten nur vor Ort und unter Aufsicht möglich. Der Aufsatz von Bird (s.o.) enthält ebenfalls einige Überlegungen dazu.

zu 4): Auch wenn das aus mir unerfindlichen Gründen viel seltener zu hören ist: Die Veröffentlichung von Transkriptionen/Aufnahmen kann selbstverständlich auch einen Vorteil im wissenschaftlichen Wettbewerb darstellen. Die ursprüngliche Anfrage von Frau Albert hat m.E. genau darauf abgezielt: "Datenarmut" ist wohl unbestreitbar ein weitverbreitetes Problem in der empirischen Sprachforschung, und angesichts des enormen finanziellen und personellen Aufwandes, den die Erstellung von Aufnahmen und zugehörigen Transkripten bedeutet, wäre es doch mit Sicherheit ein Vorteil, wenn sich größere Mengen solcher Daten wiederverwenden und gemeinsam auswerten ließen. Damit hat vielleicht nicht der Einzelne, aber mit Sicherheit die wissenschaftliche Community eher einen WettbewerbsVORTEIL gegenüber solchen Kollegen, die nicht über Transkriptsammlungen verfügen. Davon abgesehen wäre es natürlich wünschenswert, dass die Erstellung von Transkriptionen etc. als eigenständige (wissenschaftliche) Leistung anerkannt wird. Die CODATA-Initiative hat in diesem Zusammenhang ein "Konzept zur Zitierfähigkeit wissenschaftlicher Primärdaten" entwickelt (http://www.dl-forum.de/Foren/Langzeitverfuegbarkeit/Powerpointpraesentationen/PPT_Lautenschlager.ppt)[...], das sich - auch wenn es zunächst anhand meteorologischer Daten entwickelt wurde - m.E. auch für linguistische Daten als brauchbar erweisen könnte.

Entschuldigen Sie die etwas längliche e-mail, über weitere Diskussionsbeiträge würde ich mich sehr freuen.