

EINE TRANSITORISCHE PRAXIS

Spracherkennungssoftware als Reflexionsfläche akademischer Textproduktion

Ich mache das eigentlich nicht mehr

Vor zwei Jahren habe ich Kolleg_innen von meiner damaligen Gewohnheit erzählt, meine Beiträge als Vorträge zu entwerfen, die Rede vor Publikum aufzunehmen, sie automatisch zu verschriftlichen, und das Resultat als Steinbruch für Publikationen zu verwenden. Es kam der Vorschlag auf, einen Beitrag zur Rubrik «Werkzeuge» in der *Zeitschrift für Medienwissenschaft* zu schreiben, in dem ich darüber nachdenke, wie sich solche Medienwechsel auf Konzepte von mündlicher und schriftlicher Kommunikation in akademischen Textproduktionen auswirken und welche Medialitäten wir uns für unsere Vorträge und Publikationen vorstellen. Und nun schreibe ich diesen Beitrag endlich. Aber den vorliegenden Text habe ich nicht zuvor gesprochen, ich habe ihn nie als Vortrag entworfen. Das digitale Blatt unter dieser Zeile ist leer, es stehen dort keine Bruchstücke verschriftlichter Rede bereit, die ich bearbeiten könnte. Stattdessen stehen auf dem Tablet, das neben meiner Tastatur liegt, fünf handschriftliche Zeilen mit einer groben Gliederung. Meine Sätze entstehen zwischen vorübergehenden halblauten Lippenbewegungen und klappernden Tasten, wie früher schon einmal: Subvokalisierung und Tippen. Vor 30 Jahren habe ich so ähnlich geschrieben: Zwar standen der handschriftliche Entwurf und die getippten Zeilen auf Papier statt auf elektronischen Bildschirmen – auf dem Schreibtisch lag das eine Blatt und das andere war ganz zu Anfang noch in einer alten Schreibmaschine eingespannt, ja, wirklich –, aber eben beides schriftlich. Der Anteil der gesprochenen Sprache an meinem Schreiben ist wieder flüchtiger geworden.

Vorübergehend schien mir das Verfahren mit der Spracherkennung eine Entelechie, in der sichtbar und zugleich klar wurde, was ich tun will, wenn ich wissenschaftliche Beiträge verfasse. So klar ist das jetzt nicht mehr.

Verfahren

Voraus ging die Erkenntnis, dass Diktatsoftware erheblich leistungsstärker ist, als ich zuvor meinte. Wir stehen mit der automatischen Sprach- und Schrifterkennung, der automatischen Übersetzung, der Sprachsteuerung und der semantischen Sprachverarbeitung in der Mitte einer computerlinguistischen Revolution, in der unzählige kleine funktionale Verbesserungen einen Umschlag aus einer Quantität von technischen Innovationen in die Qualität tatsächlich einsetzbarer Werkzeuge vollziehen. Kulturell scheinen wir diese Revolution noch kaum vollzogen und medienhistorisch bemerkt zu haben. Wir erzählen einander noch immer die Geschichten vom Scheitern dieser Programme: die absurde Übersetzung, der mechanische Satzbau, die ungelenke Lexik. Haben wir darüber übersehen, dass sich der Kontext dieser Misserfolge verschoben hat? Ist der mechanische Satzbau in der Übersetzung des Fachartikels aus dem Chinesischen nicht heute verständlich, wo er früher opak war? Warum ist dennoch die schlechte Wortwahl erzählenswerter als die doch viel erheblichere Tatsache, dass ich, keines Wortes Chinesisch mächtig, chinesische Beiträge lese? Ist die absurde Transkription nicht inzwischen durchschaubar falsch, behebbbar und also brauchbar? Sind die Ausreißer in der Lexik heute nicht Gestalten vor dem Hintergrund des sonst richtig erkannten Textes, statt nur vor der Folie eines imaginierten Universalübersetzers? Grund für einen Verdacht: Eine ganze Phalanx neuer Technologien wird gegenwärtig uneigentlich eingesetzt. Wir verkennen sie: Als was sie verkauft werden, warum wir sie einsetzen und wie wir sie einsetzen, sind drei verschiedene Dinge.

Mich trieb etwa kein Vertrauen, sondern erst eine verletzte Hand dazu, als Notlösung und eher widerwillig *automatische Spracherkennung* einzuführen. Der unumstrittene Marktführer «Dragon Naturally Speaking» war im Sommer nach jener unglücklichen Bergtour 2010 schon hervorragend und ist seither immer besser geworden. Er stellt sich auf die eigene Stimme ein, wenn man sich die Mühe macht, ihn eine Viertelstunde ernsthaft zu trainieren. Er stellt sich ebenso auf den eigenen Sprachgebrauch ein, auch in Mischungen aus mehreren Sprachen, wenn man ihm eine große Anzahl von fertigen Texten in Textverarbeitungsdateien und E-Mails bereitstellt, aus denen er ein persönliches Profil konstruiert. Nuscheln irritiert ihn erstaunlich wenig, und fast sicher unbekannte ebenso wie frei erfundene Wörter von «Anti-Ikonizität» bis «Amene-Horas» errät er erstaunlich gut.

Das gilt immer noch. Googles aktuelle Alternative, heute im Kern der gleiche Algorithmus kombiniert mit tiefen neuronalen Netzen, ist inzwischen auch erfolgreich, personalisiert jedoch in viel geringerem Maße und ist nach meiner Erfahrung heillos überfordert, wenn er Texten oder gar Sätzen begegnet, die mehrere Sprachen mischen. Auch mein Google-Assistent weiß zwar gespens-tisch genau, wann ich gegen Abend gerne Sushi esse, aber nicht, dass ich auch

englische Sätze ernst meine und «What is the weather like today?» beantworten, nicht übersetzen oder als Phrase im Internet suchen lassen will. Fremdwörter wie «ikonisch», «konisch», «tonisch» oder auch nur «ironisch» findet er ausnahmslos «komisch».

Ich kenne eine Handvoll Kolleg_innen, die ebenfalls regelmäßig mit Spracherkennungssoftware arbeiten: Niemand unter ihnen erzählt, sie oder er habe sich gezielt dafür entschieden. Das Verfahren schleicht sich aus irgendeinem Anlass ein, es diszipliniert zum Umgang mit der neuen Schnittstelle und remediatisiert sie später, sodass, was als neue Oberfläche zum alten Schreiben hinzutrat, später als dessen sinngebender Inhalt erscheint. Als mein ursprünglicher Anlass für den Einsatz der Software entfiel und die Funktionslust andauerte, begann ich E-Mails und Posts wieder zu tippen, aber Vorträge spracherkannt zu verschriftlichen.

Ohne die Vorträge selbst nochmals hören zu müssen, konnte ich nun eine Audiodatei in eine Worddatei verwandeln. Letztere war ein Steinbruch, in drei Hinsichten von einem publizierfertigen Text entfernt: weil ich unfertig formuliere, weil die Software mich unvollständig verstand und weil Vorträge keine Aufsätze sind. Aber die ersten beiden Arbeiten am Medienwechsel waren geeignet, die unabschließbare dritte Arbeit zu kaschieren.

Übergänge

Über anderes bin ich inzwischen unsicher. *Warum* lohnte sich dieser andere Umgang mit der Mündlichkeit, die zuvor mit dem Vortrag endete und nun Material für weitere Arbeitsschritte wird? Lässt sie sich allein über den Abschied vom alten Fetisch des Mündlichen erklären?

Es schien: wegen der Modifikationen, die es mir an Rede und Schrift nahelegte. Sprechen und Schreiben waren in eine Serie von Übergängen eingespannt, eine andere als zuvor, und diese schien besser. Noch nie wollte ich Vorträge vorlesen, weil ich keine schriftlichen Texte hören und meinen schriftlichen Texten niemand zuhören wollte. Lesen wollte sie aber auch keiner. Die mündlich fundierten dagegen schon: Meine Beiträge, mündlich wie schriftlich, waren demnach vorher allzu schriftlich gewesen, schien es. Man hätte meinen können, die Aussicht auf die Verschriftlichung hätte den mündlichen Vortrag in seiner Konzeption schriftlicher werden lassen. Das Gegenteil war der Fall. Der Augenblick der kreativen Formulierung war ins Mündliche verlegt; in der schriftlichen Arbeit fand dann Überarbeitung, aber keine Schöpfung mehr statt. An das Publikum wurde im Schreiben mit jeder wiedererkannten spracherkannten Formulierung erinnert. Für gegenwärtiges Hören und spätere Schrift zu sprechen hieß, mündlicher schreiben, nicht schriftlicher reden.

Der Vergleich zwischen den beiden Modellen der Spracherkennung von Dragon Naturally Speaking und Google kann das vielleicht teilweise illustrieren.

Dragon suggeriert erstens, dass es von einigen vorgelesenen Beispielsätzen und von allen meinen schriftlichen Texten auf meiner Festplatte dasselbe lernt: Wie ich spreche, d. h., wie ich schreibe. Dragon nimmt eine Audiodatei an, ohne dass sie für Menschen, also über Lautsprecher und in verständlicher Geschwindigkeit, jemals wieder abgespielt werden müsste. Kurz: Dragon lässt sich diktieren. Es hört zweitens, wenn auch zeitversetzt und gerade deshalb konzeptionell schon schriftgemäß, zu, was ich sage; ich muss es nicht mehr hören. Und es denkt über das, was es gehört hat, im Kontext meiner Schriften nach. Es konkurriert deshalb drittens nicht mit dem menschlichen mündlichen Publikum.

Google will dagegen, dass ich mit ihm spreche, und zwar jetzt und hier. (Und es ist ein «es», keine virtuelle Person mit Namen wie Siri oder Cortana.) Was schiefgeht, liegt an missverstandenen Situationen: Ich wollte doch gerade das Wetter erfahren, ich wollte doch diesmal eine Übersetzung, ich wollte doch hier eine Suche. Was GPS und WLAN *gerade jetzt* über mich verraten, ist Google wichtiger, als wie ich *immer* mein Rolle. Wenn ich mit anderen Menschen rede («O.K. Google können wir also abschreiben»), unterbricht mich der Assistent, weil er meint, sein Name wäre gefallen: «O.K. Google!» Die Wörter hat er verstanden, den Sprechakt nicht. Er könnte das aufschreiben, aber er will es beantworten. Dragon kennt nur einen Sprechakt: das Diktat. Google diszipliniert dahingehend, meinem Vokabular von ihm leicht verstandene Befehle zugrunde zu legen. Kannst du mich dazu bringen, etwas Nützliches zu tun – findest du das Zauberwort? Dragon akzeptierte mein Vokabular und disziplinierte mich dazu, zu sprechen, wenn ich schreiben wollte: Probier mich aus, was kannst du mit mir alles schreiben? Ich finde alle deine Wörter!

Dass meine Vorträge mündlich waren, passte Dragons Verschriftlichung zwar gut; aber inzwischen hat sich die mediale Konstellation erneut verschoben. Erstens: Je visueller meine Vorträge werden, desto mehr entwerfe ich sie weder mündlich noch schriftlich, sondern in Folien, und desto lieber schreibe ich später einen ersten Entwurf an ihnen entlang. Symptom (wer tut das noch?) ist die unsichtbare Folie, die das Publikum nie sieht, sondern die mir als Notiz gilt, und die ich dennoch als Folie anlege statt als Kommentar. Und zweitens: Vorträge und Artikel verhalten sich nun seltener wie mündliche und schriftliche Fassung zueinander. Artikel und Kapitel korrespondieren vielen verschiedenen Vorträgen und Diskussionen. Das andere war ein Verfahren für Tagungsbände. Aber auch für Tagungsbände will ich heute keine Tagungsbandbeiträge mehr schreiben.

Reflexionen

Welche Gattung schreibe ich da überhaupt? Das mit dem Drachen mache ich also nur noch selten, wie ich gemerkt habe, als ich nun darüber schreiben wollte. Umso deutlicher ist mir jetzt, dass ich nicht mehr weiß, welche

Medialität welchem Beitrag gerecht wird. Das reiht sich in eine Serie von Zweifeln über die angenommene Kommunikationssituation ein: Mit welcher Absicht publizieren wir in einer Disziplin, die keine zielführenden umfassenden Bibliografien kennt, sodass unsere Publikation eher durch Netzwerke als durch Themensuchen gefunden werden wird? Zu wem reden wir, wenn wir vor einem Publikum reden, das immer noch nicht sicher weiß, ob wir ihm etwas vorlesen werden oder nicht? Mit welcher Erwartung können Tagungsveranstalter_innen, aber auch Zeitschriftenherausgeber_innen nach einem Vortrag nach *der* Publikation fragen? Ich hatte eine Zeitlang ein Verfahren, das dazu Antworten anbot. Die sind heute falsch.
