



Beat Siebenhaar (Lausanne und Bern)

## Sprachsynthese als Methode für die Dialektologie

Hier soll ein Blick auf eine wissenschaftliche Methode gegeben werden, deren Grundzüge erst erarbeitet werden müssen. Es geht darum, die Sprachsynthese nicht nur für ihren eigentlichen Zweck, Texte vorzulesen, zu verwenden, sondern auch als Instrument für die Dialektologie und generell für die Sprachwissenschaft. Nach einem Blick auf bestehende dialektologische Fragen und Methoden zeige ich Forschungslücken auf und biete Hinweise, wie die Synthese weiterführen kann. Dabei sollen Möglichkeiten aufgezeigt werden, ohne dass schon konkrete Ergebnisse vorliegen. Die erwähnten Tonbeispiele sind herunterzuladen von <http://www.unil.ch/imm/docs/LAIP/wav.files/SiebMeth.zip>.

Ein ganz besonderer Dank gebührt Eric Keller und Brigitte Zellner Keller für wertvolle Anregungen, Hinweise, Motivation und Unterstützung.

### 1 Methoden und Ergebnisse der Dialektologie

Die Dialektologie hat – wie andere Zweige der Linguistik – ein Problem nach dem andern analysiert und ist mit immer feineren Methoden zu immer differenzierteren Ergebnissen gekommen. Dabei sind einerseits viele Ortsmundarten beschrieben worden, wobei diese Monographien meist nicht die vollständige Grammatik und Lexik darstellen, sondern nur Teilbereiche. Andererseits haben wir Überblicke zu einigen grammatischen Aspekten über ein ganzes Dialektgebiet wie Entwicklung und Stand der Vokalsysteme in den Schweizer Mundarten (Haas 1978), Umlautprobleme im Schweizerdeutschen (Lüssy 1974) oder zu einzelnen Wortfeldern, wie Weinbauterminologie (Egli 1982) oder traditionelle Schifffahrt (Bickel 1995). Zum Dritten finden wir im Gefolge der pragmatischen Wende eine Vielzahl von Arbeiten, die einzelne Aspekte in sozialem Kontext ansehen: Die Vokalisierung des L in Knutwil (Christen 1988) oder R im Schweizerdeutschen (Werlen 1980), Sprachvariation und Sprachwandel (u. a. Wolfensberger 1967, Hofer 1997, Siebenhaar 2000) oder Sprachvariation im situativen Kontext (Bürkli 1999).

Schon diese exemplarische Aufzählung zeigt, dass zu vielen Details der Mundartgrammatik umfangreiche Beschreibungen vorliegen. Vor allem die segmentale Phonetik,

Phonologie und die Morphologie sind relativ gut erforscht. Es gibt zahlreiche Wörterbücher. Im Bereich der Syntax ist vor allem die Wortstellung im Verbalbereich untersucht (Christen 1993, Lötscher 1997) oder dann spezifische Aspekte im Rahmen der Generativen Grammatik (Penner 1995). Hingegen gibt es bis vor wenigen Jahren kaum Arbeiten zur Phrasierung, zum Timing und zur Intonation – zur Prosodie<sup>1</sup>.

### 2 Forschungslücke Prosodie

Die frühe Intonationsforschung – in der schweizerischen Dialektologie einzelne BSG-Bände<sup>2</sup> – hat Sprachmelodie mittels musikalischer Notation beschrieben. Für die Sprachdidaktik wurden Melodieverläufe mittels Kurven dargestellt. Erst mit der Entwicklung der apparativen Phonetik wurde eine genauere Beschreibung der Intonation überhaupt möglich. Und erst dann konnten Theorien zur Intonation entwickelt werden.

Da für die Intonationsforschung also vorerst sowohl die technischen Mittel, als auch die Theorien fehlten, so trifft diese Forschungslücke nicht nur die Mundarten, sondern etwas weniger ausgeprägt auch die Standardsprachen. Die in der Standardsprache festgestellte große Variabilität hat dazu geführt, dass man auf areallinguistische Variation kaum aufmerksam geworden ist. Mit Untersuchungen zu Sprache in Interaktion (u. a. Selting 1995) und dem DFG-Projekt "Untersuchungen zur Struktur und Funktion regional-spezifischer Intonationsverläufe im Deutschen" ändert sich die Situation. Dabei zeigt sich, dass in Einklang mit der gesamten Prosodieforschung auch im Deutschen Aspekte der Sprachmelodie in Zentrum stehen, während Phrasierung und Timing weniger interessieren.

### 3 Interaktion

All diese Ansätze zeigen, dass der wesentliche Teil der Forschung je einzelne Aspekte der Sprache, bzw. der Mundart analysiert. Weit weniger werden Interaktionen zwischen den einzelnen Teilsystemen untersucht. Klar, es gibt für die Erklärung von Wandel in der Morphologie Hinweise auf veränderte Phonemsysteme und umgekehrt, diese Bezüge sind jedoch selten systematisch. Und auch innerhalb der Prosodieforschung wird relativ wenig an Interaktionen gearbeitet. Dabei ist es von der Musik her einsichtig, dass eine andere

<sup>1</sup> Timing betrifft die Dauer eines einzelnen Segments, ist also im eigentlichen Sinne nicht suprasegmental. Da die Länge von Segmenten jedoch wesentlich durch die phonotaktische Umgebung, Silbenstruktur, Position in der Silbe, im Wort, in der Phrase bestimmt ist, ist diese Zuordnung jedoch angemessen.

Phrasierung, ein anderes Tempo, eine andere Lautstärke bei derselben Tonfolge einen anderen Eindruck hinterlässt.

#### 4 Sprachsynthese als Forschungsinstrument

Wenn man von "Sprachsynthese" spricht, so herrscht noch die Vorstellung einer kalten, abweisenden Stimme mit monotoner Intonation ohne Sinnzusammenhang vor. Auch wenn die synthetische Sprache immer noch ohne Weiteres von einer natürlichen, menschlichen Sprache unterschieden werden kann, so trifft diese negative Wertung je länger je weniger zu. Die Sprachsynthese hat in den letzten Jahren wesentliche Fortschritte gemacht und klingt heute schon viel realistischer als dies noch vor Kurzem der Fall war. Damit ergeben sich völlig neue Anwendungen für Geistes- und Sozialwissenschaften.

Ich möchte hier einen neuen Zugang für die Sprachwissenschaft andeuten, der sich hinwendet zur Untersuchung von Interaktionen zwischen einzelnen Bereichen. Dazu kann die Sprachsynthese neue Wege für die Forschung aufzeigen, indem nicht mehr nur isolierte Bereiche analysiert werden können, sondern indem die gesamte Sprache als Modell erstellt werden muss. Das heißt also, dass die verschiedenen Analysen zu einem Ganzen zusammengeführt werden müssen. Dies erlaubt, das bisher in einzelnen Bereichen erarbeitete Wissen aufzunehmen und die Relevanz jedes einzelnen dieser Bereiche in der Interaktion zu verifizieren.

Bisher wurde die Sprachsynthese nicht für solche wissenschaftlichen Experimente genutzt. In einem reduzierten Umfang ist es aber schon möglich, die existierenden Text-to-Speech-Synthesen für solche Versuche zu verwenden und Manipulationen mit diesen Systemen durchzuführen. Wie Keller und Zellner Keller (i.Dr.) gezeigt haben, lässt sich relativ einfach mit Rhythmus, mit Phrasengrenzen und Grundfrequenzverläufen experimentieren. Ich möchte einige dieser Experimente mit deutschen Beispielen hier zeigen.

Die mit LAIPTTS<sub>D</sub><sup>3</sup> erzeugten Hörbeispiele<sup>4</sup> demonstrieren, wie wichtig einzelne prosodische Aspekte sind, um Sprache zu verstehen, und wie bedeutsam die Interaktionen

<sup>2</sup> u.a. Vetsch 1910, Wipf 1910, Schmid 1915, Clauss 1927.

<sup>3</sup> LAIPTTS<sub>D</sub>: Laboratoire d'Analyse Informatique de la Parole Text to Speech Deutsch. Die Synthese ist zu finden unter: [http://www.unil.ch/imm/docs/LAIP/LAIPTTS\\_D\\_SpeechMill\\_dl\\_dt.htm](http://www.unil.ch/imm/docs/LAIP/LAIPTTS_D_SpeechMill_dl_dt.htm)

zwischen diesen Bereichen sind. Auch wenn diese Tondokumente nicht dialektaler sondern standardsprachlicher Natur sind (weil keine Mundartsynthesen existieren), zeigen sie klar, wo Dinge passieren, die bisher nicht beachtet wurden. Das heißt auch, dass ich auf einzelne interessante Aspekte hinweisen kann, wobei vielmehr Fragen gestellt als Antworten gegeben werden.

#### 4.1 Phrasierung

Zuerst soll die Phrasierung, die Strukturierung der Wörter im Syntagma, beachtet werden. *Stellen# Sie# sich# eine# Sprache# vor# die# zusammenhängende# Wörter# nicht# prosodisch# zusammenfasst# sondern# jedes# Wort# einzeln# realisiert* (Bsp01). Die abgehackte, unzusammenhängende Rede zeigt deutlich, dass die Strukturierung der Wörter im Satz interessieren sollte. Erst mit einer korrekten Phrasierung wird ein komplexerer Text überhaupt verständlich (Bsp02). Bis ein solches Modell steht, ist aber Verschiedenes zu berücksichtigen. Insbesondere sind zwei Aspekte zu erwähnen: *Eine Sprachsynthese muss ein Modell für die Generierung von Phrasengrenzen in längeren Sätzen beinhalten, um dadurch zwischen einzelnen Syntagmen oder Teilsätzen Pausen setzen zu können* (Bsp03). Ein Text ohne solche Grenzen, atemlos zu Ende gehetzt, ist schwer verstehbar. Kurze Sätze sind noch einsichtig, sobald aber größere Einheiten zusammenkommen, müssen Ideen zur Textstrukturierung entwickelt werden. Für das Setzen dieser Grenzen müssen Regeln aufgestellt werden, denn: *Stellen Sie# sich eine# Sprache vor, die ein# falsches Modell für# die Generierung von# Phrasengrenzen beinhaltet und, dadurch schwer# verständlich# wirkt* (Bsp04).

Phrasierung ist ein Bereich, den die germanistische Linguistik bisher weitgehend ausgeklammert hat. Untersuchungen finden sich in der Phonetik und in der Sprachsynthese. Für die Erklärung der Phrasierung stehen sich Modelle gegenüber, die sich einerseits wesentlich an der Syntax orientieren, andererseits solche, die eine psycholinguistische Motivation haben (vgl. Zellner 1994, Keller und Zellner Keller 1996). Alle Modelle erklären weitestgehend die Standardsprache. Vergleiche zwischen Sprachen lassen aber darauf schließen, dass hier universale Kriterien gelten könnten. Beispielsweise wurde im Laboratoire d'Analyse Informatique de la Parole der Universität Lausanne ein

<sup>4</sup> Die Tonbeispiele finden sich auf <http://www.unil.ch/imm/docs/LAIP/wav.files/SiebMeth.zip>

psycholinguistisch motiviertes Phrasierungsmodell für das Französische entwickelt, das auf Untersuchungen zur englischen und französischen Lesesprache (Grosjean & Collins 1979) basiert. Dieses Modell konnte mit wenigen Modifikationen auf das Deutsche übertragen werden. Neue empirische Daten aus den Mundarten und Untersuchungen spontaner Umgangssprache können die Modelle präzisieren, andererseits können mit bestehenden Modellen auch Unterschiede zwischen den Mundarten erfasst werden. Ein Bereich, der die Dialektologie bisher noch nicht interessiert hat.

#### 4.2 Timing

Als zweiter Aspekt wird das Timing beachtet, die Dauer der einzelnen Laute. Eine korrekte zeitliche Steuerung bestimmt Verständnis und Höreindruck wesentlich. In natürlicher Sprache werden die Laute verschieden lang realisiert. *Stellen Sie sich eine Sprache vor, in der alle Laute gleich lang sind* (Bsp05); eine unnatürliche Realisierung. Doch auch eine Berücksichtigung der intrinsischen Dauer und der Beeinflussung durch die Nachbarlaute sowie der Silbenstruktur genügt Ansprüchen an die Natürlichkeit nicht. *Stellen Sie sich eine Sprache vor, die kein Modell für die zeitliche Steuerung in der Umgebung der Phrasengrenzen aufweist. Sie hören, dass die Laute vor den Pausen zu kurz sind, weil unter anderem ein sogenanntes 'final lengthening' nicht eingebaut ist. Pausen werden so relativ schlecht wahrgenommen, die Sprache wirkt abgehackt* (Bsp06). Erst die Berücksichtigung der Position in der Phrase vermittelt einen natürlicheren Eindruck (Bsp07). Es ist jedoch zu vermuten, dass weitere Aspekte zuzuziehen sind<sup>5</sup>.

Diese Beispiele verdeutlichen, dass Veränderungen in einem Bereich, der die Dialektologie bisher nicht interessiert hat, wesentlich sind für das Verständnis von Sprache und den Eindruck, den hinterlässt. So lassen sich auch im Timingbereich Fragen stellen, die sich variationslinguistisch auswerten lassen. Werden Pausen in verschiedenen Dialekten gleich oder unterschiedlich markiert? Wie lange ist ein Laut im einen Dialekt, wie lange im anderen? Sind Dauerunterschiede relevant für perzeptive Unterscheidung von Mundarten?

<sup>5</sup> Die Darstellung der für das Schweizerhochdeutsch als relevant beurteilten Faktoren finden sich in Siebenhaar, Keller, Zellner Keller (2001).

#### 4.3 Intonation

Die Intonation baut auf dem temporalen Gerüste auf, das durch das Phrasen- und das Timingmodell gegeben ist. Das ist begründet durch die Annahme, dass auch der natürliche Sprecher die zeitliche Steuerung vor der Steuerung der Intonation vornimmt. *Ein Text ohne Intonation ist zwar schon verständlich, wirkt jedoch noch sehr monoton und dadurch unnatürlich* (Bsp08). Auf dieses zeitlich definierte Gerüst wird die Intonation gesetzt. Wir gehen in unserer Konzeption der Intonation von zwei unabhängigen Komponenten aus, ein Modell das von Fujisaki für das Japanische entwickelt worden ist<sup>6</sup>. Die Intonation stellt sich demnach zusammen aus einerseits Pitchbewegungen in einzelnen akzentuierbaren Wörtern, d.h. relativ schnelle Tonhöhenbewegungen, die in Bezug zur akzentuierten Silbe stehen, andererseits von Komponenten, die sich über ganze Phrasen bewegen. Beide Komponenten werden unabhängig voneinander gesteuert und für die definitive Intonationskurve superponiert. (Bsp09 mit konstanter Grundfrequenz. Bsp10 nur mit Akzentkommandos, Bsp11 nur mit Phrasenkommandos, Bsp12 mit kompletter Intonation). Diese Beispiele vermitteln deutlich, dass Sprache ohne Melodie keinesfalls natürlich klingen kann und dass die Melodie einen großen Einfluss auf die Verständlichkeit, aber auch auf die Wahrnehmung von Rhythmus und Sprachgeschwindigkeit hat. Gleichzeitig möchte ich betonen, dass für die Melodie einer Sprache nicht nur der Tonhöhenverlauf von Bedeutung ist, sondern mindestens ebenso die zeitliche Strukturierung. Wieder lassen sich Fragen stellen, die für die Dialektologie relevant sind: Was kennzeichnet 'normale' bzw. funktionsspezifische Intonationsverläufe in verschiedenen Mundarten aus? Wie unterscheiden sich verschiedene Varietäten? Wie relevant sind Unterschiede?

#### 4.4 Möglichkeiten der Sprachsynthese

Ein Mensch kann die einzelnen prosodischen Bereiche nicht völlig unabhängig voneinander verändern, weil wir den Stimmapparat nicht so präzise aktiv kontrollieren können. Eine Sprachsynthese kann dagegen den je einzelnen interessierenden Aspekt der prosodischen Struktur isolieren und so dessen Bedeutung auf den akustischen Gesamteindruck illustrieren kann. Die Sprachsynthese kann damit wertvolle Hilfe leisten, wenn es darum geht, Stimuli für Perzeptionstests zu generieren.

<sup>6</sup> Eine gut verständliche Darstellung mit einer Umsetzung ins Deutsche findet sich bei Mixdorff (1998).

Im Fall einer kompletten Synthese zeigen sich falsche Annahmen in einzelnen Teilen oder falsche Zusammenhänge, indem ein falscher oder zumindest auffälliger Output generiert wird, der hörbar und damit unmittelbar erfahrbar wird. Die folgenden Beispiele desselben Satzes, der von verschiedenen deutschen Text-to-Speech-Sprachsynthesen generiert wurde, können das illustrieren.

*Ich stehe in der Küche, wo sich Arbeit und Vergnügen, sinnliche Lust mit bewusstem Kalkül verbinden* (Bsp13: Bell Labs; Bsp14: IKP; Bsp15: TU-Dresden; Bsp16:IMS; Bsp17: LAIP)<sup>7</sup>. Diese verschiedenen Synthesen deuten auf verschiedene Modelle hin. Die der Synthese zugrundeliegenden Hypothesen lassen sich also hörbar vergleichen. An dieser Stelle soll nicht eine Kritik an den Systemen angebracht werden; es soll nur aufmerksam gemacht werden auf unterschiedliche Modellierung des Timings in Phrasengrenzen, die Grundfrequenzsteuerung in Nebensilben, die Intonationskurve bei weiterweisenden Breaks, die unterschiedliche Realisierungen von *stehe*.

Ein falscher Output kann über eine neue Hypothesenbildung, die Umarbeitung des Modells und die Überprüfung anhand des neuen Outputs korrigiert werden. Die Synthese ist also eine Methode, Hypothesen zu verifizieren. Damit ist die Sprachsynthese ein probates Mittel, um das in den Details immer komplexere und differenziertere Wissen über die Sprache direkt hörbar und überprüfbar zu machen.

#### 4.5 Grenzen der Sprachsynthese

Die Beispiele illustrieren gleichzeitig, dass die Sprachsynthese von einem natürlichen Sprecher noch weit entfernt ist. Bislang fehlen den synthetischen Stimmen die Ausdruckskraft. Menschliche Emotionen, wie Trauer, Überraschung, Ärger oder Freude können nicht ausgedrückt werden, denn die Stimmqualität der einzelnen Stimme kann mit der heute üblichen konkatenativen Signalgeneration nicht verändert werden. Gleichzeitig ist die Anzahl der Stimmen noch sehr beschränkt, da die Schaffung neuer Stimmen mit einem großen Aufwand verbunden ist.

<sup>7</sup> Die Beispiele wurden online synthetisiert unter:  
<http://www.bell-labs.com/project/tts/german.html>  
<http://www.ikp.uni-bonn.de/dt/forsch/phonetik/hadifix/Hadiq.html>

Die Beispiele zeigen auch, dass noch keine Mundartsynthese vorhanden ist. Die bisherigen Synthesen generieren die Standardsprachen, evtl. regionale Varianten der Standardsprachen. Ebenfalls existiert keine Synthese informeller Sprache und keine Synthese sozialer Varianten.

#### 5 Ausblick

Dieser Einblick in Möglichkeiten und Grenzen der Sprachsynthese weist auch auf aktuelle Aufgaben hin: Neben der Verbesserung der bisherigen prosodischen Modelle und der Schaffung von Modellen für stilistische Varianten wird an einer neuen Signalgeneration gearbeitet, welche die Qualität der Stimme wesentlich verbessern soll und die Stimmqualität flexibler modifizieren kann. Wir können davon ausgehen, dass in wenigen Jahren eine neue Stufe der Realitätsnähe und der Ausdruckskraft der Synthesestimmen erreicht sein wird.

Für die Dialektologie interessant ist ein vom Schweizerischen Nationalfonds finanziertes Projekt zur Synthese zweier mundartlicher Varianten, welche als Grundlage für die Erforschung der Prosodie dienen können. Die Synthese zweier Mundarten erlaubt, einzelne Parameter unabhängig voneinander zu modifizieren. Dadurch wird es möglich, Modelle für einzelne Teilsysteme zu kombinieren. So lässt sich der Frage nachgehen, wie stark die Intonation die Zuordnung von Mundarten bestimmt. Stellen Sie sich einen Text in Zürcher Mundart vor, der konsistent in Berner Intonation wiedergegeben wird. Wird er als Zürcher oder als Berner Mundart erkannt, oder ergibt das eine Mundart, die sprachgeographisch zwischen den beiden Regionen liegt? Wie hört sich ein Text an, der zwar Zürcher Lautung und Timing beinhaltet, aber eine Berner Melodie? Bisher stellen sich also erstmals Fragen, die Antworten sind noch ausstehend.

#### Literatur

Bickel, Hans (1995): Traditionelle Schifffahrt auf den Gewässern der deutschen Schweiz. Wort und Sache nach den Materialien des Sprachatlases der deutschen Schweiz. Aarau, Frankfurt am Main, Salzburg (Reihe Sprachlandschaft 17).

- Bürkli, Beatrice (1999): Sprachvariation in einem Großbetrieb. Eine individuenzentrierte Analyse anhand sprachlicher Tagesläufe. Tübingen/Basel (Stadtsprache - Sprachen in der Stadt am Beispiel Basels 1 / Basler Studien zur deutschen Sprache und Literatur 73).
- Christen, Helen (1988): Sprachliche Variation in der deutschsprachigen Schweiz. Dargestellt am Beispiel der I-Vokalisierung in der Gemeinde Knutwil und in der Stadt Luzern. Stuttgart (ZDL Beihefte 58).
- Clauss, Walter (1927): Die Mundart von Uri. Laut- und Flexionslehre. Frauenfeld (BSG 17).
- Egli, Alfred (1982): Weinbau im Deutschwallis. Sachkultur, Wortschatz, Sprachgeographie. Frauenfeld (BSM 23).
- Grosjean, F., & Collins, M. (1979). Breathing, pausing, and reading. *Phonetica*, 36, pp. 98–114.
- Haas, Walter (1978): Sprachwandel und Sprachgeographie. Untersuchung zur Struktur der Dialektverschiedenheit am Beispiel der schweizerdeutschen Vokalsysteme. Wiesbaden (ZDL Beihefte 30).
- Hofer, Lorenz (1997): Sprachwandel im städtischen Dialektrepertoire. Eine variations-linguistische Untersuchung am Beispiel des Baseldeutschen. Tübingen/Basel (Basler Studien zur deutschen Sprache und Literatur 72).
- Keller, Eric & Zellner, Brigitte (1996): A timing model for fast French. *York Papers in Linguistics* 17. pp. 53–75
- Keller, Eric & Zellner-Keller, Brigitte (i. Dr.): Speech Synthesis in Language Learning: Challenges and Opportunities. InSTIL 2000 Conference. Dundee, Scotland, August 2000. [http://www.unil.ch/imm/docs/LAIP/LAIPPTS\\_challenges.htm](http://www.unil.ch/imm/docs/LAIP/LAIPPTS_challenges.htm)
- Lötscher, Andreas (1997): 'Guet, sind Si doo'. Verbstellungsprobleme bei Ergänzungssätzen im Schweizerdeutschen. In: Ruoff, Arno und Löffelad, Peter (Hrsg.): Syntax und Stilistik der Alltagssprache. Beiträge der 12. Arbeitstagung zur alemannischen Dialektologie. Tübingen (*Idiomatica* 18), S. 85–95.
- Mixdorff, Hansjörg (1998): Intonation Patterns of German - Model-based Quantitative Analysis and Synthesis of F0 contours. [http://www.tfh-berlin.de/usr1/doz/mixdorff/public\\_html/thesis.htm](http://www.tfh-berlin.de/usr1/doz/mixdorff/public_html/thesis.htm)
- Lüssy, Heinrich (1974): Umlautprobleme im Schweizerdeutschen. Untersuchungen an der Gegenwartssprache. Frauenfeld (BSM 20).

- Penner, Zvi (Hrsg.) (1995): *Topics in Swiss German Syntax*. Bern, Berlin, Frankfurt am Main, New York, Paris, Wien.
- Schmid, Karl (1915): Die Mundart des Amtes Entlebuch im Kanton Luzern. Frauenfeld (BSG 7).
- Selting, Margret (1995): Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation. Tübingen (*Linguistische Arbeiten* 329).
- Siebenhaar, Beat (2000): Sprachvariation, Sprachwandel und Einstellung. Der Dialekt der Stadt Aarau in der Labilitätszone zwischen Zürcher und Berner Mundartraum. Stuttgart (ZDL Beihefte 108).
- Siebenhaar, Beat; Zellner Keller, Brigitte & Keller, Eric (2001): Phonetic and Timing Considerations in a Swiss High German TTS System. In: Eric Keller, Gérard Bailly, Alex Monaghan, Jacques Terken, & Mark Huckvale (eds.): *Improvements in Speech Synthesis*. Chichester. pp. 165–175.
- Vetsch, Jakob (1910): Die Laute der Appenzeller Mundarten. Frauenfeld (BSG I).
- Werlen, Iwar (1980): R im Schweizerdeutschen. In: *Zeitschrift für Dialektologie und Linguistik* 47, S. 52–76.
- Wipf, Elisa (1910): Die Mundart von Visperterminen im Wallis. Frauenfeld (BSG II).
- Wolfensberger, Heinz (1967): Mundartwandel im 20. Jahrhundert. Dargestellt an Ausschnitten aus dem Sprachleben der Gemeinde Stäfa. Frauenfeld (BSM 14).
- Zellner, Brigitte (1994): Pauses and the temporal structure of speech. In: Eric Keller (ed.) *Fundamentals of speech synthesis and speech recognition*. Chichester. pp. 41–62.