

6. Daten, Datenerhebung und Transkription

Die hier analysierten Interaktionsfragmente stammen aus 16 Unterrichtslektionen, die Teil eines Korpus von Deutsch L2-Lektionen aus der französischsprachigen Schweiz sind, das ich zwischen November 2008 und Juni 2010 erstellte. Die hier verwendeten Lektionen dauerten unterschiedlich lange, meistens zwischen 30 und 45 Minuten; zwei Werkstattlektionen erstreckten sich über gut 100 Minuten. Die Vor- und Nachnamen der Lehrkräfte und die Nachnamen der Schülerinnen und Schüler wurden entfernt. Die Lehrkräfte wurden durchnummieriert und als Lehrer, Lehrerin 1-4 bezeichnet; die Vornamen der Schülerinnen und Schüler wurden pseudonymisiert und durch vergleichbare, gleichsilbige Vornamen aus den gleichen Sprachen ersetzt.

Daten sind nicht einfach da, sondern werden qua Datenerhebung erst konstituiert. Zwischen die Interaktionen und deren Analyse tritt ein Abstand, der in mindestens zweifacher Weise medial vermittelt ist: Erstens müssen Interaktionen registriert, aufgenommen, gespeichert werden. Werden diese Aufnahmen anschließend transkribiert, dann geht zusätzlich ein Vielfaches an Informationspotential verloren, selbst wenn ein multimodales Transkriptionssystem verwendet wird und Fotografien beigefügt werden, um Sachverhalte zu illustrieren. Doch stellt nicht nur die mediale Vermittlung der Interaktionen für die Forschung eine Herausforderung dar, sondern auch die Notwendigkeit der Aufnahme der Interaktionen. Wie wir spätestens seit Labovs methodologischen Überlegungen zum Beobachtungsparadoxon wissen, kann die Präsenz des Beobachters und seiner Geräte die gesamte Interaktion und Sprachwahl nachhaltig beeinflussen (Labov 1972). Im Folgenden wird zuerst das Beobachtungsdispositiv, mit dessen Hilfe die Daten erhoben wurden, detailliert beschrieben und begründet. Dann wird erläutert, welche Transkriptionskonventionen verwendet wurden und weshalb. Das Beobachtungsparadoxon wird in Kapitel 7 diskutiert.

6.1 Das Beobachtungsdispositiv

Daten werden mit dem Ziel erhoben, die Interaktionen so detailliert wie möglich zu registrieren. Alle irgendwie an einer Aktivität Beteiligten und ihre interaktionalen Räume sollten erfasst, die Aktivität möglichst in ihrer Gesamtheit aufgezeichnet werden. Aufnehmen ist aber, wie Mondada im Anschluss an Macbeth formuliert, »an embodied exercise of inquiry and analysis« (Mondada 2012, 51). Die Wahl der Aufnahmegeräte und die Handhabung der Geräte, das heißt, wo und wie sie platziert werden, entscheiden darüber, welche Daten entstehen können. Platzierung der Kameras, Wahl

der Ausschnitte und der Perspektiven, Wahl des Beginns und des Endes der Aufnahmetätigkeit sind Resultate der Analyse der Situation und basieren auf Erwartungen, welche der Forschende in die Situation hereinträgt (Mondada 2012, 56). In der Aufnahmesituation selbst müssen Erwartungen, Aufnahmematerial und aktuelle Situation vermittelt werden, und die Praxis des Aufnehmenden muss allenfalls angepasst werden (Mondada 2012, 54-55). Ich filmte mit zwei Videokameras. Die beiden Videokameras wurden während des Plenumsunterrichts auf Stativen fix an zwei Standorten so platziert, dass ein möglichst großer Bereich des Klassenzimmers gefilmt werden konnte (siehe als Beispiele Abb. 6.1 und 6.2).

Abbildung 6.1



Abbildung 6.2



Dieses Vorgehen erlaubte es mir, während des Aufnehmens die Handlungsabläufe im Klassenzimmer zu beobachten und Feldnotizen zu erheben. Gleichzeitig war es das angemessenste Vorgehen, um einen möglichst guten Überblick über die Aktivitäten im Klassenzimmer zu erhalten. Zudem ist die Aufnahmetechnik auch am unaufrüttlichsten, weil die Kamera nicht ständig präsent gemacht wird (Heath, Hindmarsh & Luff 2010, 40). Zwar ist es mit zwei Videokameras nicht möglich, alle Details aller Interaktionen, die sich gleichzeitig im Klassenzimmer ereignen, zu erfassen. Doch liefern die fix installierten Kameras zumindest konsistente Ansichten gesamter Handlungsabläufe derjenigen Teile des Klassenzimmers, die sie erfassen, und sie ermöglichen ebenfalls eine umfassendere Analyse der Partizipationsstrukturen, die im Klassenzimmer sehr schnell wechseln können. Ein weiterer Vorteil dieses Vorgehens ist es, dass der Beobachter nicht zum Kameramann wird, der das Aufnehmen ständig den Handlungen anpassen muss und damit so ebenfalls ständig mit den im Klassenzimmer Handelnden interagiert, gleichzeitig aber immer Gefahr läuft, zu spät zu kommen, weil er die Handlung ja nicht nur einfangen, sondern vielmehr antizipieren muss, wenn er Handlungsabläufe in ihrer Gesamtheit erfassen will. Das Antizipieren von Handlungen dürfte nämlich eher als Glücks- denn Normalfall einzustufen sein. Macbeth (1999) und mit ihm Laurier und Philo (2012) und Heath et al. (2010) weisen nämlich darauf hin, dass die Initiierung der Kamerabewegung hin zur Handlung *de facto* oft erst durch den Beginn der Handlung ausgelöst wird und die Kamera damit den Beginn der Handlung schon verpasst hat.

Die Platzierung der Kameras musste den jeweiligen Unterrichtssituationen und Unterrichtsformen angepasst werden. Das didaktische Szenario des Werkstattunterrichts etwa, in dem die Kinder an unterschiedlichen Arbeitsstationen gleichzeitig Aufgaben erledigen, konnte mit nur zwei Kameras nicht in seiner Gesamtheit erfasst werden. Hier wurden die Kameras bei zwei Stationen platziert (Beispiele in Abb. 6.3

und 6.4). Da die Kinder rotierten, war es so möglich, Daten unterschiedlicher Gruppen zu erheben.

Abbildung 6.3

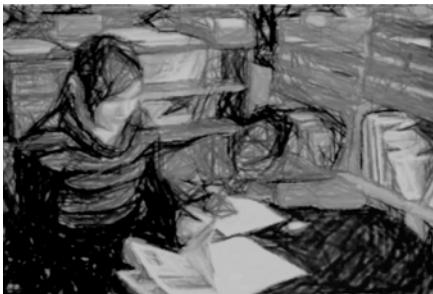


Abbildung 6.4



Der filmische Beobachtungsapparat wurde durch einen umfassenden »Belauschungsapparat« ergänzt. Um »backstage interactions«, wie man sie im Anschluss an Goffman (1959) und mit de Fina (2014) nennen könnte, oder besser einfach die Peer-Interaktionen während des Plenarunterrichts sowie die Kommunikation während Gruppenarbeitsphasen in den Pulten mitverfolgen zu können, wurden Audioaufnahmegeräte auf den Pulten der Schülerinnen und Schüler verteilt, und zwar so, dass in jeder Pultgruppe mindestens ein Audioaufnahmegerät vorhanden war. Dies wurde als Variante der innerhalb der Soziolinguistik oft praktizierten Methode gewählt, nach welcher Teilnehmende mit Audioaufnahmegeräten ausgestattet werden, die sie selbstständig bedienen können (etwa Rampton 2006, 32; Holmes & Hazen 2013, 181). Wie in Kapitel 7 gezeigt werden wird, werden die sich in Seh- und Griffweite der Kinder befindenden Audio-geräte immer wieder thematisiert, womit die Aufnahmesituation relevant wird.

6.2 Transkriptionskonventionen

Die Daten wurden im Großen und Ganzen im Anschluss an GAT 2.0 transkribiert (Selting et al. 2009). GAT 2.0 hat sich im deutschen Sprachraum in der scientific community derjenigen Gesprächsforschenden durchgesetzt, die sich methodisch zwar an der Konversationsanalyse orientieren, aber eine explizitere und einfacher zu handhabende Transkriptionskonvention wünschen als die Jefferson'sche, die in der angloamerikanischen Konversationsanalyse verwendet wird (Dittmar 2009, 141). GAT, so die Entwickler, versucht zwar, so viele Prinzipien und Konventionen der Jefferson'schen Transkription wie möglich zu übernehmen (Selting et al. 2009, 353). Gleichzeitig sind nach GAT 2.0 transkribierte Interaktionen auch für Nicht-Linguistinnen und Nicht-Linguisten relativ einfach lesbar (Selting et al. 2009, 360-361).

Als »formbezogenes Transkriptionssystem« (Selting 2001, 1061) intendiert GAT 2.0, die Organisationsstruktur der Interaktion sichtbar zu machen. Gleichzeitig ermöglicht es GAT 2.0 auch, multimodale Aspekte in die Transkription zu integrieren (Selting et al. 2009, 368-369). In dieser Hinsicht ist GAT 2.0 relativ wenig explizit. Laut GAT 2.0 werden parasprachliche und außersprachliche Handlungen, die während

eines Sprecherbeitrags vorkommen, mit Hilfe doppelter Klammern transkribiert und mit Hilfe spitzer Klammern in den gesprochenen Text integriert:

```

1 A:      ich hab <<weinend> das gar nicht so gemeint>
2 B:      ((seufzt))
3      (.)
4 B:      <<sich erhebend> aber du hast es> so gesagt (Selting et al.
2009, 368)

```

Diese Konvention erweist sich bei der Integration umfangreicherer multimodaler Information als impraktikabel. Anstelle der direkten Integration paralleler parasprachlicher und außersprachlicher Handlungen in den gesprochenen Text mit Hilfe spitzer Klammern werden Handlungen, die gleichzeitig mit einem sprachlichen Beitrag vorkommen, auf einer eigenen Zeile transkribiert, die nicht eigens nummeriert wird. Beginn und Ende werden hier mit Hilfe von senkrechten Strichen im gesprochenen Text und auf der Zeile der parasprachlichen und außersprachlichen Handlungen notiert. Außersprachliche Handlungen, die nicht zeitgleich mit sprachlichen Handlungen vorkommen, werden auf einer eigenen Zeile transkribiert und erhalten eine eigene Zeilennummer (siehe dazu auch etwa Heller & Morek 2015).

2M6-P 1

```

1 Le4:      ZAHLEN.
2      | (2.2) was ↑IST zahlen.      |
3      | ((dreht den kopf nach li)) |
4      | ((dreht den kopf nach re)) |
4      | (1.5) ein | zwei drei vier [fünf sechs sieben;] |
4      | ((bewegt re hand auf und ab))      |
5 Jda:          [((hebt die hand)) ]
6 Kor:          [((hebt die hand)) ]
7 Ros:          [((hebt die hand)) ]

```

Außersprachliche Handlungen, die andauern, werden mit Hilfe von Pfeilen transkribiert (2M1G5-20, Zeile 1). Das Ende dieser Handlungen fällt mit dem Einsetzen der nächsten transkribierten Handlung des gleichen Akteurs, der gleichen Akteurin zusammen.

2M1G5-20

```

1 Nes:      ((geht von de re seite des tisches zur li seite->))
2 Akt:      heute [fehlt eu::h      ]
3 Nes:          [((setzt sich)) ]
4 Akt:      euhm (3.0) |<<:-> ma(h)e)mi, >      |
4 Akt:          | ((schlägt mit der re hand auf den tisch)) |

```

Wie in GAT 2.0 werden aber nur dann auch para- und außersprachliche Aspekte transkribiert, wenn sie relevant erscheinen. Diese Transkriptionskonvention unterscheidet sich von der multimodalen Transkriptionskonvention, wie sie von Mondada (siehe etwa 2011; 2012) entwickelt wurde und wie sie inzwischen innerhalb der konversationsanalytisch ausgerichteten multimodalen Interaktionsanalyse schon fast zum Standard geworden ist. Doch lässt sich der von Selting et al. (2009) gegen die Jefferson'sche Transkription vorgebrachte Einwand der schwierigeren Zugänglichkeit auch gegen Mondadas Konvention ins Feld führen. Diese setzt ein beträchtliches Maß an Transkriptionslesekompetenz voraus, an deren Aneignung stärker praktisch und didaktisch orientierte Leserinnen und Leser höchstwahrscheinlich wenig interessiert sein dürften. Doch sollen die Transkriptionen und die Analysen hier auch für dieje-

nigen, deren Lektüre didaktisch und weniger konversations- oder interaktionsanalytisch motiviert sein dürfte, einigermaßen einfach lesbar sein.

Im Zusammenhang mit Daten, die aus einem französischsprachigen Kontext stammen, ergeben sich mit GAT Segmentierungsprobleme. GAT 2.0 segmentiert die einzelnen Turns der Sprechenden weiter in kleinere Einheiten. Die von Sacks et al. (1974, 702 et passim) eingeführte »Turn-Construction Unit« (TCU) ist gemäß GAT 2.0 nur bedingt als Segmentierungseinheit verwendbar. Zwar sei es durchaus möglich, dass komplett syntaktische Einheiten im Zusammenspiel mit kompletten Intonationsverläufen Einheiten konstituierten, die als semantisch mögliche Brocken (»chunks«) und als adressatenorientiert gestaltete Informationseinheiten betrachtet werden könnten. Denn TCUs seien ja auch durch das Zusammenspiel von Syntax und Prosodie bestimmt (Selting 2000, 504). Doch seien Syntax und Prosodie nicht immer synchron. In bestimmten Kontexten würden Intonationseinheiten gebildet, die nicht als TCUs interpretiert werden könnten, weil sie syntaktisch nicht mögliche vollständige Konstruktionen seien. Gleichzeitig müssten TCUs keineswegs immer mit einem »transition relevant place« TRP enden. Dass am Ende einer TCU kein TRP folgen müsse, könne mit Hilfe lokal-prosodischer, syntaktischer, lexiko-semantischer, pragmatischer Mittel projiziert werden (Selting 2000, 504-511).

In GAT 2.0 werden Turns nach Intonationsphrasen (IP) segmentiert. IP werden in GAT 2.0 im Anschluss unter anderem an Fox (2000), aber etwa auch Cruttenden (1997) definiert als eine »durch einen als kohäsig wahrgenommenen Tonhöhenverlauf« gestaltete Einheit, die »obligatorisch mindestens eine akzentuierte Silbe« enthält, also eine Silbe, »die durch eine Intonationsbewegung und/oder Lautstärke und/oder Länge phonetisch hervorgehoben wird und die semantische Bedeutung der Äußerung maßgeblich bestimmt«. Diesen nennen die Autorinnen und Autoren von GAT 2.0 Fokusakzent. Diese Intonationseinheit werde durch eine Reihe fakultativer Grenzsignale abgegrenzt: Tonhöhensprünge nach unten oder oben am Einheitsanfang, schnellere Sprechgeschwindigkeit in den Vorlaufsilen der Einheit, finale Dehnung am Ende der Einheit, Glottalisierung am Ende der Einheit, Tonhöhenbewegungen am Einheitenende auf unbetonten Silben und Pausen (Selting et al. 2009, 370).

Intonationsphrasen als Segmentierungsgrößen zu verwenden ist indessen auch nicht unproblematisch. Wie Auer festhält, wird üblicherweise gefordert, dass eine IP einen nuklearen (Fokus)-Akzent enthält. Da dies aber noch keine hinreichende Bedingung für die Identifizierung einer IP ist, werden phrasenbegrenzende Merkmale, darunter »finale Dehnung, initiales Accelerando auf dem Vorlauf, finale Intonationsbewegung (fallender oder steigender Grenzton), Reset der Deklination oder/und des pitch range zu Beginn zur Segmentierung in Intonationsphrasen herangezogen«. Letztlich bietet dieser Katalog von Merkmalen aber keine Definition, sondern lediglich eine »Heuristik für das Erkennen von IP-Grenzen«; seine Anwendung sei schwierig und führe nicht immer zu eindeutigen Ergebnissen (Auer 2010, 8). Teilweise lassen sich Intonationsphrasen sehr unterschiedlich einteilen (siehe auch Barth-Weingarten 2011).

Hier stellt sich ein weiteres Problem. Das hier verwendete Korpus besteht nicht aus englischem und auch nicht nur aus deutschem Sprachmaterial, sondern enthält viel französisches Sprachmaterial. Fürs Deutsche mag die Definition des Hauptakzents als Fokusakzent, welcher die Bedeutung der Informationseinheit maßgeblich mitstrukturiert, adäquat sein. Fürs Französische kann der Fokusakzent nicht als notwendiges Kriterium einer IP definiert werden, da das Französische die Intonation gänzlich

anders als das Deutsche organisiert (Hirst 1998). Im Französischen operieren erstens gleichzeitig zwei unterschiedliche Akzentsysteme. Der »interne Akzent« (Jin 1990, 18ff.) oder, wie er auch genannt wird, der »finale Akzent« (Blanche-Benveniste 2005, 162) liegt auf der letzten Silbe einer akzentuierbaren Einheit. Er hat eine deliminative Funktion. Diesen Akzent hat Cruttenden im Blick, wenn er darauf hinweist, dass das Französische den Nukleus der Intonationsphrase nicht im gleichen Maß verschieben könne wie etwa das Englische, und, könnte man ergänzen, das Deutsche. Deswegen sind die Intonationseinheiten, er nennt sie Intonationsgruppen, normalerweise kürzer als in den anderen indoeuropäischen Sprachen, deren Prosodie heute bekannt ist:

French appears to make more use of the options involving short intonation-groups than does English. This is related to the fact that French does not have the mobile nucleus characteristic of English [...] nor the same potential of pre-nuclear accents, and hence is forced to introduce extra intonation-groups for highlighting. (Cruttenden 1997, 140)

Wenn man sich an diesen Intonationsgruppen für die Transkription orientieren würde, dann würden die Zäsurierungen innerhalb eines Turns ziemlich kurz:

2m1-P 1

- 1 Aln: monSIEUR?
 2 euh y'a une camerA?
 hat es eine kamera
 3 là-D'SUS?
 da drauf

In gewissen Kontexten taucht der externe Akzent (Jin 1990, 18) oder, von Blanche-Benveniste so genannt, »l'accent d'instance« (Blanche-Benveniste 2005, 164) auf. Der externe Akzent fällt auf den Anfang einer akzentuierbaren Einheit oder auf ein Wort, welches normalerweise nicht akzentuiert ist. Dieser Akzent hat eine rhematisierende Funktion und kommt dann zum Einsatz, wenn aufgrund syntaktischer Stellungsbeschränkungen andere rhematische Markierungen nicht möglich wären (Jin 1990, 26). Zwar lassen sich die beiden Akzentsysteme unterscheiden, da der Vokal beim externen Akzent nicht gelängt werden kann. Typischerweise wird er »durch einen abrupten Fo-Sprung um ein bis zwei Niveau« realisiert (Jin 1990, 24; siehe auch Blanche-Benveniste 2005, 164-165). Da der externe Akzent aber keineswegs obligatorisch ist, kommt er auch nicht in allen Intonationseinheiten vor; die beiden Systeme überlagern sich, wie in folgendem Ausschnitt:

2m6-P 37

- 1 Le4: c'est un VERbe; (1.0)
 das ist ein verb
 2 jean-da.
 3 donc il faut JUSTE que vous REMETTIEZ dans vos têtes en
 sixIÈME;
 also ist es nur nötig dass ihr in der sechsten
 von neuem daran denkt
 4 (1.5) que quand on vous demande la signification d'un
 VERbe.
 dass wenn man euch nach der bedeutung eines verbs fragt
 5 il y aura probablement une signification en franÇAIS,
 es auf französisch wahrscheinlich eine bedeutung geben wird
 6 qui sera AUSSI un vErbe.
 die auch ein verb sein wird

Dazu kommt ein weiterer Umstand in meinen Daten: Das Deutsch, das in meinem Korpus verwendet wird, ist eine L2-Varietät des Deutschen und weist teilweise spezifische prosodische Eigenschaften auf. Das wird etwa in folgendem Turn hörbar:

2m1-P 713

1 Le4: sie KOMMT aus der TÜRkei.

In diesem Turn kommen zwei Akzente vor. Der zweite Akzent scheint deplatziert zu sein, da er auf der ersten Silbe von TÜRkei liegt. Intonationsphrasen im Sinne von GAT 2.0 zu verwenden, ist also ebenfalls problematisch.

Gibt es denn ein sinnvolleres Vorgehen des Segmentierens? Dem Segmentieren von Turns in kleinere Einheiten haftet immer etwas Problematisches an, denn Segmentierungen, die von Forschenden ja *ex post* vorgenommen werden, laufen im Prinzip Gefahr, etische Manöver zu sein, die über die Köpfe der in die Interaktion involvierten Teilnehmenden durchgeführt werden. Ein Linguist, der segmentiert, ist in einer grundsätzlich anderen Situation als eine Interagierende, die im Turn ihres Gegenübers laufend darüber entscheiden muss, wann nun ein möglicher Abschlusspunkt im Beitrag des Gegenübers erreicht sein könnte. Auer greift aus diesem Grund auf eine Konzeption der Segmentierung oder eben Zäsurierung zurück, die gerade diese Aufgabe der in die Interaktion Involvierten zum Ausgangspunkt macht. Diese Aufgabe erklärt er phänomenologisch-wahrnehmungspsychologisch:

Teilnehmer an direkten Interaktionssituationen sind hingegen in die zeitliche Emergenz der Interaktion gezwungen; es gibt für sie kein Hin- und Herspringen zwischen später und früher. Sie müssen, um überhaupt an der Interaktion teilnehmen zu können, das Gehörte sofort danach beurteilen, ob es sich zu einer abgeschlossenen Einheit zusammenfügt oder als offene Gestalt noch einer Fortführung bedarf. Einmal getroffene Entscheidungen können revidiert werden; so erweist sich, was vielleicht zuerst als vollständig und abgeschlossen erschien, oft im nächsten Augenblick als Bruchstück einer größeren Gestalt. (Auer 2010, 11)

Optimale Gestaltschlüsse sind nach Auer dann erreicht, »wenn an einem Punkt sämtliche syntaktische, prosodische und semanto-pragmatische Projektionen abgearbeitet sind« (Auer 2010, 11-12). Daneben gibt es Gestaltschlüsse, bei welchen nur einige dieser Merkmale auftauchen, und natürlich auch uneindeutige Fälle. Auer schlägt nun ein Zäsurierungssystem vor, in welchem Zäsuren in der Syntax, Zäsuren in der Prosodie und semantisch-pragmatische Zäsuren markiert werden.

Dieses Vorgehen ist sicherlich dann wichtig, wenn genau diese Zäsuren im Zentrum des Interesses stehen, was hier aber nicht der Fall ist. Die Idee, auf Gestaltschlüsse zu achten und nicht auf spezifische, *a priori* festgelegte Merkmale, wird aber auch für die Transkription hier übernommen. Die Einteilungen werden aber nicht spezifisch markiert.

