



HAL
open science

Zur Transkription von Lautobjekten in der Gesprächsanalyse

Heike Baldauf-Quilliatre

► **To cite this version:**

Heike Baldauf-Quilliatre. Zur Transkription von Lautobjekten in der Gesprächsanalyse. Cordula Schwarze; Sven Grawunder. Transkription und Annotation gesprochener Sprache und multimodaler Interaktion. Konzepte, Probleme, Lösungen, Narr, pp.74-100, 2022, 9783823384694. hal-03942176

HAL Id: hal-03942176

<https://hal.science/hal-03942176v1>

Submitted on 16 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Heike Baldauf-Quilliatre

Zur Transkription von *Lautobjekten* in der Gesprächsanalyse

Abstract:

Lautobjekte (auch bezeichnet als *vocalizations* oder *liminal signs*) stellen die gesprächsanalytische Transkription vor eine Herausforderung, da sie nicht oder nur bedingt orthographisch transkribiert werden können. Neuere Untersuchungen haben aber gezeigt, dass eine genaue Transkription auch für ein besseres Verständnis der Interaktionsorganisation wesentlich ist. Dieser Artikel plädiert daher in Bezug auf Lautobjekte für eine systematische Integration phonetischer Transkription in ein gesprächsanalytisches Basistranskript. Er bezieht sich dabei sowohl auf herkömmliche Praktiken der Transkription von Lautobjekten in der Konversationsanalyse, als auch auf eine Beispielanalyse des Lautobjekts „pff“.

Keywords:

Gesprächs-/Konversationsanalyse, non-lexical vocalizations, sound objects, Interjektion, phonetische Transkription

1. Einleitung

Wer mündliche Interaktion transkribiert, trifft früher oder später auf sie: Laute, die keinem Wort im traditionellen Sinn entsprechen, die nicht oder nicht vollständig lexikalisiert sind und denen die Beteiligten dennoch eine Funktion zuschreiben. Die grammatische Kategorie der Interjektion suggeriert eine Einheitlichkeit und Abgeschlossenheit, die sich bei näherer Betrachtung schnell als trügerisch erweist¹ – auch in Bezug auf die Möglichkeit einer schriftlichen Darstellung. Manche Lautverbindungen gelten gemeinhin als Interjektionen und sind unter Umständen sogar in Wörterbüchern zu finden (z.B. *igitt*, *au*, *brr*). Sie können in der Regel relativ problemlos orthographisch transkribiert werden - dabei fasst man jedoch unter Umständen lautliche Realisierungen, die verschieden verwendet werden, unter einer (Schrift)Form zusammen. Für andere Laute bzw. Lautverbindungen gibt es mehr oder weniger konventionalisierte Schreibungen aufgrund ihrer Verwendungen in bestimmten

¹ Darauf weisen grammatisch orientierte Arbeiten allerdings auch immer wieder hin (z.B. Nübling 2004; Ehlich 2007).

schriftlichen Kommunikationsformen bzw. Textsorten (z.B. *oh*, *hm*). Das stellt im Alltag kein Problem dar, ist aber in einer gesprächsanalytischen Transkription zum Teil zu ungenau und zieht im Weiteren Schwierigkeiten in der Analyse nach sich, etwa in Bezug auf die temporale Entfaltung der Interaktion. Wieder andere können phonetisch beschrieben und transkribiert werden (z.B. Clicks). Ihre Transkription mit Zeichen des lateinischen Alphabets ist aber problematisch, da eine genaue gesprächsanalytische Transkription relativ schnell unlesbar (und nicht nachvollziehbar) wird bzw. die Buchstaben Laute suggerieren, die dem eigentlich realisierten Laut nur bedingt entsprechen. Und schließlich gibt es jene Laute, die auch das Internationale Phonetische Alphabet vor Herausforderungen stellen (z.B. Ächz- und Stöhnlaute, wie sie Hofstetter (2020) analysiert).

Seit Beginn gesprächs- und konversationsanalytischer Forschungen wird darauf hingewiesen, dass diese Lautobjekte, ebenso wie Schweigephasen bzw. Pausen, Wiederholungen oder andere sogenannte syntaktische „Disfluenzen“, nicht ignoriert werden dürfen (siehe etwa Schegloff (1982), Goodwin (1986)). Im Gegenteil, sie sind wichtige Elemente in der Interaktionskonstitution und tragen wesentlich zur Organisation (aus Teilnehmer*innensicht) und zum Verständnis der Interaktion (aus Forscher*innensicht) bei. Auch wenn die lautliche Form an sich in gesprächsanalytischen Arbeiten nicht im Vordergrund steht, stellt sich daher durchaus die Frage nach der Darstellung dieser Laute bzw. Lautverbindungen in der Transkription.

Im Folgenden (Abschnitt 2) soll zunächst aufgezeigt werden, wie Lautobjekte in Gesprächs- und Konversationsanalyse konzipiert sind. Diese Frage ist insofern wichtig, als sie Überlegungen zur Transkription in bestimmte Richtungen lenkt (lenken kann). Im Anschluss daran (Abschnitt 3) werden mit Bezug auf gesprächs- bzw. konversationsanalytische Arbeiten Fragen und Probleme hinsichtlich ihrer Transkription diskutiert; es geht also in gewisser Hinsicht um Fragen der Phonem-Graphem-Beziehung aus gesprächsanalytischer Sicht. Ein dritter Teil (Abschnitt 4) argumentiert an Hand eines nicht-lexikalisierten Lautobjekts wie „pff“, warum es sinnvoll ist, zumindest Nicht-lexikalisiertes in gesprächsanalytischen Arbeiten phonetisch zu transkribieren.

2. Von *response cries*, *vocalizations*, *sound objects* und *liminal signs*

Als einer der ersten hat Erving Goffman (1981) eine Gruppe von nicht- oder halblexikalisierten „Ausrufen“ (*response cries*) und ihre Funktionen in der Interaktion beschrieben. Die Besonderheit dieser *response cries* besteht u.a. darin, dass sie keine Antworthandlung verlangen und es somit den anderen Interaktionsteilnehmer*innen erlauben, sie „zu überhören“. Andererseits handelt es sich aber um soziale Phänomene („creatures of social situations“; 1981: 121), die in eine interaktionale Sequenz integriert sein können. Goffman beschreibt *response cries* als eine spezifische Form von „Herausplatzern“ (*blurtings*). In etwa vergleichbar, aber weniger klar ist der Begriff *outloud*, den Szymanski (1999) in Bezug auf Goffmans „Selbstgespräch“ (*self talk*) benutzt: „Outloud utterances may not be designed in their production specifically to re-engage a next turn at talk; however, they can provide a source for future talk about specific aspects of the task.“ (Szymanski 1999: 19).

Konversationsanalytische Studien sprechen in Anlehnung an Schegloff (1982) weitestgehend von „Vokalisierung“ (*vocalization*). Schegloff versteht darunter zunächst einmal mehr oder weniger lexikalisierte Formen (wie ‚uh huh‘ oder ‚yeah‘), die in der Regel als *continuer*² benutzt werden und daher nicht das Rederecht einfordern (1982, 73-74). Mit Verweis auf die zunehmende Forschung in diesem Bereich unternehmen Keevallik & Ogden (2020) eine genauere Begriffsbestimmung. Sie beschreiben Vokalisierungen als „körperliche Prozesse“ (*bodily processes*; 2020: 8): „some vocal tract sounds have a more obviously somatic origin, while, like laughter and crying, they can be positioned in meaningful ways in sequences of action“ (2020: 8). Keevallik & Ogden stellen ebenfalls heraus, dass es sich um ein Kontinuum mit mehr oder weniger konventionalisierten Formen handelt, von denen manche phonetisch unterspezifiziert sind (im Sinne von Keating 1988³) bzw. genuin multimodal (das heißt sie sind eng mit bestimmten körperlichen Aktivitäten verwoben). Die Beiträge in dem von Keevallik & Ogden herausgegebenen *Special Issue* der Zeitschrift *Research on Language and Social Interaction* zeigen, wie wichtig die Berücksichtigung der Körperlichkeit des Sprechens vor allem hier ist und dass eine genaue Analyse eine genaue Transkription aller Parameter einschließt – sowohl multimodale, als auch phonetisch-prosodische.

² Siehe dazu Goodwin (1986).

³ „For the vocalization to be recognizable as a token of a particular type of vocalization, speakers have to produce some phonetic events, with freedom to vary some elements.“ (Keevallik & Ogden 2020: 6)

Im Bereich der Interaktionalen Linguistik ist ebenfalls regelmäßig die Rede von „Lautobjekten“ (*sound objects*). Dieser Begriff wurde von Reber & Couper-Kuhlen (2010) geprägt und in Reber (2012) weiterentwickelt. Den Autorinnen geht es bei der Begriffswahl darum, auf die Rolle phonetisch-prosodischer Parameter für die Bedeutung im semantischen Sinne hinzuweisen: „the label sound object aims to reflect the fact that these objects are spoken language resources for which the sound pattern and its context-specific use are distinctive for the meaning“ (Reber 2012: 12).

Ein neues und meines Erachtens sehr interessantes Konzept umfasst der von Dingemanse (2020) geprägte Begriff der „Grenzsignale“ (*liminal signs*). Dingemanse argumentiert gegen die negative und ausschließende Beschreibung dieser Lautobjekte (etwa als *nicht*-lexikalisiert oder *nonverbal*) und plädiert für eine Hervorhebung des In-between-Charakters dieser Laute oder Lautverbindungen: „Liminal signs are signs that derive interactional utility from being ambiguous with regard to conventionality, intentionality, and accountability. Their in-between status [...] is an essential part of their form and function, as they occupy the interstices of talk and frequently serve to navigate liminal and transitory spaces in interaction.“ (Dingemanse 2020: 191). Die Analyse von Grenzsignalen wird möglich durch und bedarf zum einen einer genauen Darstellung (unter anderem um die genaue zeitliche Entfaltung in der Interaktionssequenz zu verstehen) und zum anderen einer genauen theoretischen Konzeption (etwa durch die Verbindung mit anderen Disziplinen und deren Forschungsergebnissen in diesem Bereich).

Ich werde im Folgenden von *Lauten* bzw. *Lautverbindungen* sprechen, wenn es sich um eine spezifische phonetische Realisierung handelt, ohne dass eine Aussage über eventuelle Funktionen in der Interaktion getroffen wird. Um auf eine bestimmte Gestalt zu verweisen, die eine Funktion in der Interaktion ausübt, werde ich Rebers Begriff des *Lautobjekts* verwenden.

3. Lautobjekte transkribieren: Fragen, Vorschläge und Probleme

Der deutliche Bezug auf die Rolle phonetisch-prosodischer und körperlicher Aspekte für die Konstitution von Lautobjekten, vor allem in neueren interaktionsphonetischen Arbeiten, verweist auch auf Fragen zu deren Transkription. Wenn nämlich Laute oder Lautverbindungen benutzt werden, die normalerweise nicht in der untersuchten Sprache

vorkommen oder wenn es sich um Laute handelt, die eng an bestimmte körperliche Aktivitäten geknüpft sind (wie etwa das Heben eines schweren Gegenstandes), dann greift eine Transkription, die sich an der jeweiligen (Schrift)Sprache orientiert oder die körperliche Phänomene ignoriert, im besten Fall zu kurz.

Transkriptionen dienen in erster Linie dazu, die Flüchtigkeit mündlicher Äußerungen oder Interaktionen festzuhalten. Sie erlauben es im Weiteren, Daten zugänglich und Analysen für andere nachvollziehbar zu machen. Das setzt voraus, dass sie bestimmten, mehr oder weniger einheitlichen Konventionen folgen, die in einer bestimmten Wissenschaftsgemeinschaft bekannt und anerkannt sind.

Aus gesprächs- bzw. konversationsanalytischer Sicht ist Transkribieren mehr als nur ein genaues schriftliches Abbilden von Interaktionen. Transkribieren ist Teil des analytischen Prozesses (Psathas & Anderson 1990) und wird daher auch entsprechend reflektiert (z.B. ten Have 2002, Bucholtz 2007, Davidson 2010, Bolden 2015). Deppermann & Schütte (2008) weisen darauf hin, dass beim Transkribieren epistemologische und pragmatische Entscheidungen getroffen werden, etwa hinsichtlich der Granularität in Abhängigkeit vom zu untersuchenden Phänomen, von der interaktionalen Relevanz oder vom Adressaten des Transkripts (siehe auch Deppermann 2001). Ayass (2016) spricht von einer *Generierung* der Daten durch die Transkription („The process of transcription *generates* the data upon which the analysis is built.“ 2016: 511). Mondada (2007: 809) beschreibt Transkripte als „evolving flexible object“ und Transkribieren als „situated practice“.

Lautobjekte müssen in gesprächsanalytischen Transkripten immer und systematisch erfasst werden. In Bezug auf die Art der Transkription stellt sich nun aber die Frage: Wie genau soll ein Lautobjekt transkribiert werden, so dass seine phonetisch-prosodische Realisierung und seine Einbettung in den Interaktionsverlauf einerseits nachvollziehbar wird, andererseits aber nicht die Lesbarkeit des Transkripts schmälert? Und inwiefern müssen sichtbare Aspekte ihrer Realisierung (Reber & Couper-Kuhlen 2010) bzw. körperliche Aktivitäten, die damit in Zusammenhang stehen können (Keevalik & Ogden 2020) systematisch in der Transkription berücksichtigt werden?

Erste Überlegungen zur Transkription von Lautobjekten im weiteren Sinne finden sich bereits in den 1980er Jahren bei Gail Jefferson. Im Rahmen ihrer Untersuchungen zum Lachen plädiert sie gegen eine Beschreibung als „((laughter))“ und für eine genaue Transkription der Lachpartikel (Jefferson 1985). Diese ermöglichen es nämlich, die genaue Platzierung in

einem bestimmten Turnsegment zu bestimmen, was wiederum die Analyse der Lachsequenz beeinflusst (siehe auch das entsprechende Kapitel in Hepburn & Bolden 2017). Potter & Hepburn (2009) führen diese Überlegungen weiter, indem sie Lachpartikel innerhalb eines Wortes als „interpolare Aspirationspartikel“ (*interpolated particles of aspiration*) bezeichnen. Was zunächst wie eine (zu) komplizierte terminologische Wortschöpfung erscheinen mag, ist eine durchaus berechtigte Ausdifferenzierung, die auf der Annahme basiert, dass die Bezeichnung als *Lachpartikel* diese Phänomene in die Nähe des Lachens rückt und damit auch bestimmte Interpretationen nahe legt. Das im deutschen Sprachraum übliche *Gesprächsanalytische Transkriptionssystem* (Selting et al. 2009) unterscheidet sich diesbezüglich von Jeffersons Konventionen: GAT2 bietet sowohl die Beschreibung („((lacht))“), als auch eine Transkription als silbisches Lachen („hahaha“) an; sprechbegleitendes Lachen wird als „vorangestellte Beschreibung mit Angabe der Extension erfasst“, vom Einfügen von Lachpartikeln in Wörter wird abgeraten, da diese die automatische Suche in Transkriptionen beeinträchtigen (Selting et al. 2009: 367).

Klassische konversationsanalytische Studien orientieren sich in der Regel an Jefferson (2004). Lautobjekte werden dabei durch Buchstaben oder Buchstabenverbindungen transkribiert, die dem ursprünglichen Laut so nah wie möglich kommen. Charakteristische prosodisch-phonetische Aspekte wie Länge, Stimmqualität, Behauchtheit etc. sind durch entsprechende Zeichen markiert (siehe dazu Hepburn & Bolden 2017).

Für stärker konventionalisierte oder sogar lexikalisierte Formen wie „oh“, „ah“, „aha“, „hm“, „äh“ und ähnliches stellt dies insofern ein geringeres Problem dar, als es sich um eine konventionalisierte Schreibung handelt. Schwierig wird es aber für weniger konventionalisierte Formen. Hepburn & Bolden (2017: 93) geben in diesem Fall folgenden allgemeinen Ratschlag: „Generic advice in this situation would be to first have a go at the basic sound. For instance, for a throat clear one would expect an out-breath element, a guttural element, ,g', ,gh', ,ugh', etc., sometimes a ,k' sound, all of which i soften done through closed or sealed lips, giving a possible ,m' sound, so ,hgm' or ,mhgmhm'.“

Konkret etwa bietet Hepburn (2004: 262) für Schniefen in englischsprachigen Wein-Sequenzen (*sniffing*) eine Transkription mit „.shih“ (Schniefen in Verbindung mit Tränen) bzw. „.skuh“ (Schniefen in Verbindung mit einem Schnauben im Rachenraum) an – der dem „s“ vorangestellte Punkt signalisiert eine Lautproduktion während der Inspiration. In Hepburn & Bolden (2017, 83-84) erwähnt sie auch noch „.snih“. Hoey (2020) transkribiert

Schniefen ohne Weinen (*sniffing*) mit „>.nh<“ – „><“ signalisiert ein schnelleres Tempo in Bezug auf das Sprechtempo in der Umgebung. Mondada (2020) untersucht Schnüffeln in Riech-Sequenzen (*sniffing*) und transkribiert es ebenfalls als „.nh“ mit Variationen in Hinsicht auf Lautheit und Tempo. In allen drei Arbeiten wird das untersuchte Phänomen als *sniffing* bezeichnet und die Transkription zeigt, dass es sich um bei der Inspiration produzierte Laute handelt; ansonsten bestehen aber deutliche Unterschiede: So sind, wenn man sich auf die Transkription bezieht, die sogenannten „feuchten Schniefer“ nur in Hepburn’s Wein-Sequenzen zu finden. Mondada’s „Schnüffler“ in Riechsequenzen sind immer nasal, aber von verschiedener Länge. Hoey’s Schniefer, die in Konversationen spezifische Funktionen im Rahmen des Turntaking ausüben, sind ebenfalls nasal, werden aber immer mit sehr hohem Tempo produziert – was nicht der Fall ist für Mondadas Beispiele.

Diese Beobachtung führt nun zu mehreren Überlegungen: Als erstes kann man feststellen, dass diese Unterschiede bei einer Beschreibung des Lautobjekts als „((sniff))“ entgehen würden (siehe dazu auch Dingemanse 2020: 192). Nun könnte man natürlich einwenden, dass die Granularität der Transkription von der jeweiligen Fragestellung abhängt (siehe etwa Mondadas (2007) Beschreibung von Transkribieren als situierte Praktik). Dies setzt aber voraus, dass man tatsächlich genauer transkribiert, sobald das Lautobjekt in der Analyse relevant wird. Zwar ist das Verbessern und Verfeinern von Transkripten in Gesprächs- und Konversationsanalyse durchaus gängige Praxis, aber problematisch bleibt die Frage, ob eine (zu) ungenaue Transkription nicht dazu führt, dass manche Forschungsfragen gar nicht gestellt werden. Als zweites wäre zu fragen, ob hier nicht Dinge terminologisch zusammengefasst werden, die eigentlich nicht zusammen gehören. Wenn es sich um Lautphänomene handelt, die nicht nur Unterschiede hinsichtlich ihrer Realisierung, sondern auch hinsichtlich ihrer Verwendung aufweisen und wenn die Realisierungsarten mit verschiedenen Verwendungsweisen zu korrelieren scheinen, dann führt das Zusammenfassen unter einem Begriff meines Erachtens in eine falsche Richtung. Hier wird nämlich suggeriert, dass es sich um ein und dasselbe Phänomen handelt (siehe etwa die erwähnten Beispiele verschiedener Arten von „Schnüfflern“ und ihrer Funktionen). Ich möchte in diesem Zusammenhang auch noch einmal auf Potter & Hepburn (2009) verweisen, die meines Erachtens zu Recht auf den Unterschied zwischen Lachen und sogenannten *Lach*partikeln innerhalb eines Wortes hinweisen und für eine terminologische Trennung plädieren. Drittens kann man sich fragen, bis zu welchem Grad eine Transkription

funktioniert, die versucht, Laute, für die es in der untersuchten Sprache keine Entsprechung gibt, im Rahmen einem sprachgebundenen Transkriptionssystems wiederzugeben. Wäre es denn in diesem Fall nicht sinnvoller, auf andere Transkriptionssysteme, wie etwa das Internationale Phonetische Alphabet, zurückzugreifen? Das mag zwar auf den ersten Blick gewöhnungsbedürftig erscheinen, es würde aber meines Erachtens eine bessere Nachvollziehbarkeit und Genauigkeit erlauben. Und verschiedene konversationsanalytische Arbeiten zeigen, dass dies durchaus möglich und sinnvoll ist (etwa Hofstetter 2020; Keevallik & Ogden 2002; Couper-Kuhlen & Reber 2020; Wiggins & Keevallik 2020).

Im Gegensatz zu phonetischen Arbeiten geht es in Gesprächs- und Konversationsanalyse weniger um die möglichst genaue Darstellung des Lautobjekts, sondern darum zu zeigen, dass und wie es die nachfolgende Rede beeinflusst bzw. dass und wie es sich auf die vorangegangene Rede bezieht. Dies ist meines Erachtens allerdings kein Widerspruch. Wenn ein Lautobjekt in der Interaktionskonstitution eine Rolle spielt, dann betrifft dies nicht nur die sequenzielle und multimodale Einbettung, sondern ebenso die lautliche, phonetisch-prosodische Realisierung. Ungenauigkeiten oder vorschnelle Verallgemeinerungen können hier zu falschen Interpretationen führen. Dies zeigt sich zum Beispiel bei Untersuchungen zu (lexikalisierten) Diskursmarkern und Antwortpartikeln: So wurde etwa gezeigt, dass deutsches „jaja“ nicht nur in unterschiedlichen Sequenzen und sequenziellen Positionen auftritt, sondern dabei auch unterschiedlich realisiert wird (Golato & Fagyal 2008; Barth-Weingarten 2011). Für eine genaue Analyse bedarf es hier einer prosodisch genauen Transkription. In Arbeiten zu den französischen Diskursmarkern „oh“ und „ah“ wird darauf hingewiesen, dass „ah“ zwar wesentlich häufiger vorkommt und dass beide in vergleichbaren sequenziellen Kontexten auftreten können, es sich jedoch um zwei unterschiedliche Lautäußerungen handelt (Fauré & Vérine 2001; ICOR 2008). Weiterführende Untersuchungen zu „oh là là“ erwähnen mögliche Unterschiede zu ähnlichen Formen wie „ah là là“ oder „roh là là“ (Baldauf-Quilliatre et al. 2016). Auch hier wird deutlich, dass eine phonetisch genaue Transkription (im Sinne der Unterscheidung zwischen [a], [o] und [xo]) für eine genaue Analyse unabdingbar ist. Ein letztes Beispiel: In ihren Untersuchungen zum Ausdruck von negativen oder positiven Geschmackserlebnissen beschreibt Wiggins (2002) einen emphatisch, mit steigend-fallender Tonhöhenbewegung realisierten, stark gelängten Nasal [m:], der in Anwesenheit von Speisen bzw. Getränken produziert wird und dazu dient, ein positives (angenehmes) sensorielles Erlebnis zu äußern.

Dieses *gustatory mm* unterscheidet sich von anderen „mm’s“, etwas sogenannten *acknowledgement tokens* oder *continuers* (Gardner 2001) nicht nur durch die spezifische sequenzielle Platzierung, sondern auch durch eine spezifische phonetisch-prosodische Realisierung und eine spezifische Verbindung mit anderen multimodalen Ressourcen wie Blick, Mimik, etc. (Gonzalez Temer 2017; Wiggins & Keevallik 2020). Was also etwa in einer Basisversion als „mm“ transkribiert wird, können letztlich sehr unterschiedliche Phänomene sein, bei denen die gleiche Schreibweise eine Gemeinsamkeit suggeriert, die möglicherweise gar nicht existiert. Je genauer die Transkription, desto feiner lässt sich im Weiteren auch differenzieren. Das zeigen etwa die Untersuchungen von Gonzalez Temer (2017) und Wiggins & Keevallik (2020), die jeweils nicht nur die genaue prosodische Realisierung verschiedener *gustatory mm’s*, sondern auch die jeweilige materielle Realisierung bzw. die Verbindung mit verschiedenen anderen multimodalen Ressourcen transkribieren. So verbinden zum Beispiel Wiggins & Keevallik (2020) eine konversationsanalytische Transkription nach Jefferson und eine multimodale Transkription nach Mondada (2018) mit logarithmischen Tonhöhenkurven und können dadurch zeigen, dass *gustatory mm* in verschiedenen Sequenzen auftreten kann, dabei verschieden realisiert wird und verschieden in andere Handlungen der Teilnehmer*innen eingebunden ist .

All diese Beispiele machen deutlich, wie problematisch es sein kann, wenn scheinbar ähnliche (oder ähnlich klingende) Phänomene unter einer Form zusammengefasst werden. Ein genauerer Blick zeigt unter Umständen, dass diese Phänomene gar nicht so ähnlich sind. Dies betrifft nicht nur das sequenzielle Umfeld, sondern auch die genaue lautliche, also phonetisch-prosodische Realisierung und die multimodale Einbettung. Das bedeutet nun nicht, dass Lautobjekte immer systematisch phonetisch-prosodisch transkribiert werden müssen. Aber so, wie es inzwischen zur guten Forschungspraxis gehört, in einem interaktionslinguistischen Transkript Fokusakzente und Intonationsphrasen zu notieren und so, wie der multimodale Charakter der Interaktion sich auch immer stärker in Transkription und Analyse widerspiegelt, so sollte man auch darüber nachdenken, bestimmte Phänomene wie nicht-lexikalisierte Lautobjekte phonetisch-prosodisch genauer zu transkribieren – zumindest dann, wenn man sich in der Analyse darauf bezieht.

4. Plädoyer für eine phonetische Transkription: das Lautobjekt „pff“

In diesem Teil möchte ich an Hand eines konkreten Beispiels zeigen, wie wesentlich die phonetisch genaue Transkription ist, wenn es darum geht, ein Lautobjekt zu beschreiben bzw. zunächst einmal zu bestimmen. Ich beziehe mich dabei auf eine stimmlose konsonantische egressive Lautverbindung, die Schmidt et al. (2015)⁴ in ihren Konventionen zu cGAT erwähnen und als „pff“ transkribieren.

Die Annahme einer solchen Lautäußerung ist allerdings nicht selbstverständlich: In deutschen Wörterbüchern gibt es in der Regel keinen Artikel über eine Interjektion / ein Lautobjekt „pff“ und auch in Arbeiten zu deutschen Interjektionen, die grammatisch argumentieren oder auf schriftsprachlichen Korpora beruhen, wird „pff“ kaum erwähnt. Allerdings findet man Beispiele für „pff“ in Arbeiten zur Kurznachrichtenkommunikation (z.B. Balnat 2012). Außerdem gibt es (schriftsprachliche) Interjektionen, die dem nahe kommen könnten. So führt zum Beispiel Nübling (2004: 15) in ihrer Liste prototypischer Interjektionen neben „pfui!“ eine konsonantische Interjektion „phh!“ an, die Gleichgültigkeit und Verachtung ausdrücken soll und zwei weitere Interjektionen „puh!“ und „pah!“, die dazu dienen, Erleichterung und Ekel bzw. Verachtung und Geringschätzung anzuzeigen.

Betrachtet man nun Studien, die sich auf authentische gesprochensprachliche bzw. interaktionale Korpora beziehen, ist ein Lautobjekt „pff“ durchaus präsent. So beschreibt etwa Pompino-Marschall (2004) in seiner Studie eines 45-minütigen Fernsehinterviews 83 „echte“ Interjektionen und listet an dritter Stelle mit einer Frequenz von 12% eine Interjektion „pff“, die verschieden realisiert werden kann (mit entstimmlichem bilabialem Trill, stimmlosem bilabialem Frikativ und anderen Frikativvarianten; Pompino-Marschall 2004: 78) und die zwischen 110 und 675 Millisekunden dauert. Im Rahmen ihrer Konventionen für computergestütztes Transkribieren in Anlehnung an GAT2 erwähnen Schmidt et al. (2015) „pff“ als Beispiel einer Interjektion, „die nicht im Duden verzeichnet ist und für die damit auch keine empfohlene Schreibung existiert“ (Schmidt et al. 2015: 21). In diesem Fall, so die Autor*innen (siehe auch Winterscheidt et al. 2019), sollte man sich einerseits für eine Schreibweise entscheiden („pff“ an Stelle von „pf“ oder „ph“, um die Verwechslung mit einem Wortabbruch zu vermeiden), andererseits aber phonetische Unterschiede berücksichtigen (z.B. „pft“, wenn die Interjektion mit einem Plosiv endet).

⁴ Es handelt sich dabei um ein Handbuch für computergestützte Transkription (mit dem Transkriptionseditor FOLKER), mit dem im Rahmen der Datenbank FOLK gearbeitet wird. Die Transkription folgt im Wesentlichen dem gesprächsanalytischen Transkriptionssystem GAT2, weist aber einige Ergänzungen oder Änderungen auf.

Winterscheidt et al. (2019: 16) erwähnen „pff“ dabei im Rahmen der usuellen Interjektionen, auch wenn es keinen Wörterbucheintrag dafür gibt.

In allen diesen Fällen wird das Lautobjekt als „pff“ transkribiert, in Anlehnung an Jeffersons Prinzip, eine Buchstabenverbindung zu finden, die der tatsächlichen Lautverbindung möglichst nahe kommt. Auch ohne eine genauere Untersuchung stellt sich hier bereits eine erste Frage: Wenn „pff“ phonetisch verschieden realisiert wird (so wie es Pompino-Marschall suggeriert) und wenn es sich um keine lexikalisierte Form mit Wörterbucheintrag handelt, wenn es keine oder kaum eingehende Untersuchungen bzw. Beschreibungen gibt, wie kann man dann davon ausgehen, dass es sich bei all den Varianten um das selbe Lautobjekt handelt? Wie kann man bestimmte phonetische Unterschiede in der Transkription berücksichtigen – und etwa den plosiven Abschluss als eigenständige Form interpretieren, die entsprechend anders (als „pft“) transkribiert werden muss, während andere Varianten wie etwa ein bilabialer Trill an Stelle eines bilabialen Plosivs oder ein bilabialer Frikativ an Stelle eines labiodentalen unberücksichtigt bleiben?

Im Folgenden soll nun detailliert gezeigt werden, was einerseits für die Annahme eines Lautobjekts „pff“ spricht und was andererseits eine phonetische Transkription im Rahmen einer gesprächsanalytischen Beschreibung dieses Lautobjekts leisten kann. Ausgangspunkt dafür ist die Datenbank FOLK des Leibniz-Instituts für deutsche Sprache, die 376 Okkurrenzen von „pff“ und 133 Okkurrenzen von „pf“ verzeichnet (Stand: 01.03.2021). Die Häufigkeit des Auftretens rechtfertigt im Übrigen die Kategorisierung als usuelle Interjektion (Winterscheidt et al. 2019): Betrachtet man im Vergleich dazu andere usuelle Interjektionen mit Wörterbucheintrag, findet man bei einigen eine ähnliche oder sogar geringere Anzahl von Okkurrenzen (etwa „uff“ mit 327 Okkurrenzen zuzüglich 147 Fälle von „uf“ oder „tja“ mit 338 Okkurrenzen).

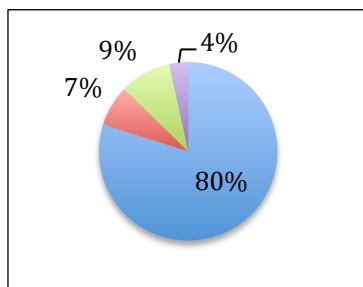
Für diese nachfolgende Untersuchung wurden alle Fälle von „pff“ auditiv phonetisch und sequenziell analysiert. Interessant ist „pff“ aus interaktionsphonetischer Sicht vor allem wegen seiner Stimmlosigkeit, die keine prosodische Analyse zulässt. Soweit möglich (nicht im Overlap, keine parasitären Geräusche, etc.) wurde die auditive Analyse durch PRAAT unterstützt. Dabei bewegt sich die phonetische Analyse hier in bestimmten Grenzen: so wurden etwa Längen nicht in Bezug gesetzt zum Sprechtempo, Intensität nicht zur durchschnittlichen Intensität des*der Sprecher*in etc. Eine solche, phonetisch genaue Analyse würde es sicher erlauben, die hier getroffenen tendenziellen Aussagen zu

präzisieren. Bei Videodaten wurden ebenfalls multimodale Aspekte, insbesondere bezüglich der materiellen Realisierung der Lautverbindung (soweit sichtbar) hinzugezogen.

In einem ersten Schritt schließt die Analyse alle Fälle aus, in denen „pff“ für mich nicht hörbar war. Eine segmentalphonetische Untersuchung der 353 „echten“ Lautverbindungen zeigte daraufhin folgendes Bild:

- hinsichtlich des ersten Segments:

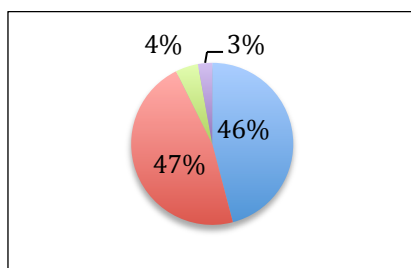
In 80% der Fälle (282 Okk.) wurde ein bilabialer stimmloser Plosiv produziert [p], in 7% der Fälle (26 Okk.) handelt es sich um einen (stimmhaften oder entstimmten) Trill/Vibranten [b] oder [β]. In 9% der Fälle (33 Okk.) war kein Plosiv oder Trill nachweisbar, die Lautverbindung bestand nur aus dem Frikativ; 4% (12 Okk.) waren unklar.



Tab. 1

- hinsichtlich des zweiten Segments:

In 47% der Fälle (165 Okk.) produzierten die Sprecher*innen einen stimmlosen labiodentalen Frikativ [f], in 46% (162 Okk.) handelt es sich um einen stimmlosen bilabialen Frikativ [ɸ], in 4% der Fälle (16 Okk.) war kein Frikativ nachweisbar und in 3% (10 Fälle) war das Segment eher unklar.



Tab 2

In der überwiegenden Mehrheit der Fälle wird also tatsächlich ein bilabialer Plosiv und ein bilabialer oder labiodentaler Frikativ realisiert, so wie es die Transkription als „pff“ suggeriert. Interessant sind aber nun meines Erachtens die abweichenden Fälle und die Grenzen. Mit anderen Worten:

- Besteht die Lautäußerung obligatorisch aus zwei Segmenten (Plosiv/Trill + Frikativ), d.h. handelt es sich bei Formen wie [p], [β], [f] oder [ϕ] um andere Lautobjekte (d.h. Lautverbindungen mit anderen Funktionen in der Interaktion)?
- Besteht ein Unterschied zwischen der Verwendung eines Plosivs und eines Trills bzw. eines bilabialen oder labiodentalen Frikativs?

Plosiv/Trill

Ein Blick auf die 16 Fälle, in denen kein Frikativ zu hören ist (obwohl die Transkription in FOLK durchgängig „pff“ aufweist), zeigt, dass es sich dabei immer um einen stimmhaften oder entstimmlichten Trill handelt. Beispiel 1 stellt einen solchen Fall dar:⁵ Es handelt sich um einen Ausschnitt aus einer Fahrschulstunde: die Fahrschülerin soll mit dem Auto rückwärts einparken und dafür die Räder einschlagen. Der Ausschnitt beginnt, wenn sie mit den Rädern an der Bordsteinkante anstößt.

Beispiel 1 FOLK_E_00167_SE_01_T_01⁶

(https://dgd.ids-mannheim.de/DGD2Web/jsp/DisplayVideoTranscriptExcerpt.jsp?v_session_id=11C016FFD1DB3268E42D97462BEA0DF8&v_doc_id=FOLK_E_00167_SE_01_T_01_DF_01&v_contribution_id=c790&v_word_id=w1431&v_context_ids=&v_context_size=4&v_text_size=200&v_download=true)

778	RK	n jetzt kanns du nich mehr lenken ne		
779		(0.24)		
780	EA	ja		
781		(0.5)		
782	RK	so un jetzt kanns du nur noch hoffen		
783		(0.2)		
784	RK	passt oder passt nich		
785		(1.3)		
786	RK	weil jetzt hab ich den fehler gemacht		
787		(0.92)		
788	RK	weil Ich bin zu SCHNELL gefahrn,		
789		(0.28)		
790	RK	Aber? #1(.)	#2 [β.]	#3 (.)

⁵ Die Transkription „pff“ in FOLK wird hier und in den folgenden Beispielen systematisch durch eine IPA-Transkription ersetzt.

⁶ Die Transkription folgt den Konventionen von GAT2 (Selting et al. 2009).



#1



#2



#3

791 EA passt gAr nicht (mehr) beSTIMMT,
792 RK das MACHT nichts jetzt [...]

Der entstimmte Trill [ɸ] ist hier deutlich sichtbar durch die aufgeblähten Wangen und vorgestülpten Lippen (#2) – im Gegensatz zu #1 und #3. Er befindet sich am Ende einer längeren Erklärsequenz, in der RK den aktuellen Zustand formuliert (Z. 778), davon ausgehend die zukünftigen Handlungen der Fahrschülerin projiziert und sie damit vor eine Entscheidungssituation stellt (Z. 782 und 784), um schließlich in einem Account die Gründe für diesen Zustand anzugeben, wobei er sich selbst als Schuldigen bezeichnet (Z. 786 und 788). Dieser relativ lange Turn wird abgeschlossen durch den Konnektor „aber“, der eine adversative oder konzessive Fortführung ankündigt und ein von Mikropausen eingerahmtes [ɸ]. Eine vergleichbare Position findet man auch bei Formen mit Frikativ. Im letzteren Fall zeigt der Sprecher jedoch an, dass a) eine bereits zuvor geäußerte Bewertung oder Schlussfolgerung bekannt ist und nicht noch einmal erklärt werden muss und dass b) eine erneute Wiederholung nicht nur unnötig, sondern auch unerwünscht ist (siehe Beispiel 5).

In Beispiel 1 ist dies jedoch nicht der Fall. Zwar interpretiert die Fahrschülerin EA den Turn ebenfalls als abgeschlossen und antwortet auf die indirekte Frage, die mit der Aufforderung zur Entscheidung verbunden war (Z. 791), aber RK führt seinen Turn in Z. 792 fort. Es handelt sich somit nicht um einen Abschluss, sondern ausschließlich um die Ankündigung einer Positionierung des Sprechers. Es geht auch nicht darum, auf etwas rückzuverweisen und es als schon gesagt und bekannt darzustellen; [ɸ] projiziert eine Positionierung des aktuellen Sprechers.

Betrachtet man nun im Weiteren die Realisierung des Plosivs/Trills in allen Okkurrenzen, fällt auf, dass dieser regelmäßig und individuell unabhängig mehr oder weniger stark aspiriert und mit einem mehr oder weniger deutlichen Verschluss produziert wird. Dabei handelt es sich wie bereits erwähnt nur um tendenzielle Angaben, da individuelle und andere Besonderheiten in der Artikulation bei der Analyse nicht berücksichtigt wurden. Ausschnitt 2 zeigt ein Beispiel mit einem intensiven Verschluss, Ausschnitt 3 einen eher schwachen.

In Beispiel 2 handelt es sich um ein Telefongespräch zwischen zwei Freundinnen.

Beispiel 2 FOLK_00084_SE_01_T_01

[https://dgd.ids-](https://dgd.ids-mannheim.de/DGD2Web/jsp/DisplayTranscriptExcerpt.jsp?v_session_id=11C016FFD1DB3268E42D97462BEA0DF8&v_doc_id=FOLK_E_00084_SE_01_T_01_DF_01&v_contribution_id=c619&v_word_id=w3450&v_context_ids=&v_context_size=4&v_text_size=200&v_download=true)

[mannheim.de/DGD2Web/jsp/DisplayTranscriptExcerpt.jsp?v_session_id=11C016FFD1DB3268E42D97462BEA0DF8&v_doc_id=FOLK_E_00084_SE_01_T_01_DF_01&v_contribution_id=c619&v_word_id=w3450&v_context_ids=&v_context_size=4&v_text_size=200&v_download=true](https://dgd.ids-mannheim.de/DGD2Web/jsp/DisplayTranscriptExcerpt.jsp?v_session_id=11C016FFD1DB3268E42D97462BEA0DF8&v_doc_id=FOLK_E_00084_SE_01_T_01_DF_01&v_contribution_id=c619&v_word_id=w3450&v_context_ids=&v_context_size=4&v_text_size=200&v_download=true)

618 FR und habt ihr schon irgendwelche plÄne fürs WOchenende,
619 EG °h ja [Bϕ] äh ja ich [xxx] heute mit [name] was MAchen,

FR stellt eine Entscheidungsfrage, die aber, im Gegensatz zum Format der Frage, im Rahmen der Interaktion einer weiteren Elaboration bedarf (Raymond 2003). EG antwortet auf diese Frage zunächst mit der Minimalantwort „ja“, erweitert ihre Antwort aber in einem nachfolgenden Account (Z. 619). Die Lautverbindung befindet sich, gemeinsam mit einer Zögerungspartikel „äh“ als eine Art Pivot zwischen der Minimalantwort und der Erklärung (Baldauf-Quilliatre 2019). In der Mehrzahl der Fälle dieser Art (Pivot zwischen präferierter Minimalantwort und Account) scheint der Verschlusslaut relativ intensiv. Aber es gibt auch Gegenbeispiele bzw. andere Sequenzen, in denen der Verschlusslaut mit relativ hoher Intensität realisiert wird.

Beispiel 3 ist ein Ausschnitt aus einem sprachbiographischen Interview, in dem der Rentner AK von zwei Studentinnen interviewt wird.

Beispiel 3 FOLK_00147_SE_01_T_02

[https://dgd.ids-](https://dgd.ids-mannheim.de/DGD2Web/jsp/DisplayTranscriptExcerpt.jsp?v_session_id=11C016FFD1DB3268E42D97462BEA0DF8&v_doc_id=FOLK_E_00147_SE_01_T_02_DF_01&v_contribution_id=c591&v_word_id=w2288&v_context_ids=&v_context_size=4&v_text_size=200&v_download=true)

[mannheim.de/DGD2Web/jsp/DisplayTranscriptExcerpt.jsp?v_session_id=11C016FFD1DB3268E42D97462BEA0DF8&v_doc_id=FOLK_E_00147_SE_01_T_02_DF_01&v_contribution_id=c591&v_word_id=w2288&v_context_ids=&v_context_size=4&v_text_size=200&v_download=true](https://dgd.ids-mannheim.de/DGD2Web/jsp/DisplayTranscriptExcerpt.jsp?v_session_id=11C016FFD1DB3268E42D97462BEA0DF8&v_doc_id=FOLK_E_00147_SE_01_T_02_DF_01&v_contribution_id=c591&v_word_id=w2288&v_context_ids=&v_context_size=4&v_text_size=200&v_download=true)

581 EH wie is denn das verhältnis zu den LEUten [...]
585 kommen die mit den TIERen zurecht und äh [wie sin]d die so
586 AK [die äh]
587 EH dr[Auf,]
588 AK [räuspert sich] (.) wir verstEhen uns eigentlich ALLE,
589 ich ich ich äh
590 (0.25)
591 AK °hhh (.) sAch ma eh
592 (0.47)
593 AK ich bin oäh [pf:]

594 (0.35)
 595 AK ja FRIEDlich; ((lacht)) °hh [also]
 596 EH [UMgä]nglich;
 597 AK äh [umgänglich] (.) ja
 598 EH [((lacht))]

In Zeile 581-587 stellt EH stellt eine Frage, auf die AK antwortet (Z. 586-597). Ab Z. 588 deuten Wiederholungen, Pausen, Reformulierungsindikatoren, Zögerungspartikel u.ä. auf Formulierungsschwierigkeiten hin (z.B. Gülich & Kotschi 1996). Er beginnt daraufhin eine neue Konstruktion („ich bin“, Z. 593), schließt aber erneut mit verschiedenen Zögerungsmerkmalen wie einer nicht usuellen Partikel, der Lautverbindung [pf:] und einer Pause an, bevor er die Konstruktion mit „ja friedlich“ (Z. 595) zu Ende führt. Wie in den meisten Fällen dieser Sequenz tritt die Lautverbindung nach der Zögerungspartikel „äh“ (bzw. ihrer Varianten, hier „oäh“) und vor lexikalisierten epistemischen Markern unterschiedlicher Art (hier die Partikel „ja“) auf. Das deutet darauf hin, dass die Lautverbindung nicht nur dazu dient, Formulierungsprobleme anzuzeigen. Der Verschlusslaut ist dabei in der Regel kurz und vor allem wenig intensiv. Allerdings gibt es auch hier Gegenbeispiele und eine vergleichbare geringe Intensität findet sich auch in anderen sequenziellen Kontexten.

Ob es sich bei starkem oder schwachem Verschluss nur um Varianten handelt oder ob die Stärke des Verschlusses auf verschiedene Funktionen hinweist, lässt sich nur durch eine Sequenzanalyse mit einer systematischen, phonetisch genauen Transkription des Lautobjekts feststellen. Mit aller Vorsicht könnte man sagen, dass in bestimmten Sequenzen möglicherweise eher ein starker Verschluss zu erwarten ist, während in anderen der Verschluss wahrscheinlich schwach wird. Dabei handelt es sich aber keinesfalls um voneinander abgegrenzte Kategorien, sondern sicherlich eher um Tendenzen, die gleichzeitig mit der Art der Positionierung, der lautlichen Umgebung, eventuellem phonetischem Alignment, und nicht zuletzt habitueller und individueller Sprechweise korrelieren. Zu einem vergleichbaren Schluss kommt auch Proske (2014) in Bezug auf „komm“, für das sie ebenfalls Tendenzen aufzeigt, ohne zum Zeitpunkt des Aufsatzes eine genauere Kategorisierung vornehmen zu können.

Frikativ

Wie stellt sich nun die Situation in Hinblick auf den frikativischen Laut dar? Es wurden 33 Fälle identifiziert, in denen nur ein bilabialer (17 Okk.) oder labiodentaler (16 Okk.) Frikativ auftritt. Auch hier wurden in FOLK alle Okkurrenzen als „pff“ transkribiert und somit dem Lautobjekt „pff“ zugeordnet. Die Frikative können dabei sowohl lang (16 Okk.), als auch kurz (17 Okk.) sein – wobei es sich auch hier um relative und approximative Angaben handelt. Die Laute werden in ähnlichen sequenziellen Positionen und mit ähnlichen Funktionen benutzt, wie die Lautverbindung mit bilabialem Plosiv bzw. Trill, auch wenn es einige Unterschiede bezüglich der Lautverbindung/des Lautes „Plosiv/Trill + Frikativ“ oder „Plosiv/Trill“ gibt. So zeigt eine genauere Untersuchung der sequenziellen Einbettung, dass es sich in der Mehrzahl um einen Laut handelt, der innerhalb eines Turns produziert wird und von Pausen oder anderen Partikeln eingerahmt ist (16 Fälle) oder der allein oder mit einem epistemischen Disclaimer einen Antwortturn darstellt (9 Fälle). Betrachtet man diese 25 Okkurrenzen aus sequenzieller und soweit möglich multimodaler Sicht, wird deutlich, dass es sich um Fälle handelt, in denen der Frikativ entweder dazu beiträgt ein Formulierungsproblem anzuzeigen (sei es als Indiz für Ungenauigkeit und Vagheit, sei es als Suche nach dem passenden Element) oder auf ein Disalignment bzw. Disengagement des Sprechers hinweist. Ausschnitt 4 zeigt einen Beitrag aus dem Saalpublikum im Rahmen einer Podiumsdiskussion.

Beispiel 4 FOLK_E_00391_SE_01_T_04⁷

https://dgd.ids-mannheim.de/DGD2Web/jsp/DisplayVideoTranscriptExcerpt.jsp?v_session_id=11C016FFD1DB3268E42D97462BEA0DF8&v_doc_id=FOLK_E_00391_SE_01_T_05_DF_01&v_contribution_id=c254&v_word_id=w3133&v_context_ids=&v_context_size=4&v_text_size=200&v_download=true

254 DH [...] i hab jetzt einfach mal GANZ a polemische frAge;

255 °h (.) ma rEd immer so von baNAnarepublik; #4 h°



#4

256 +°h (.) so+#5 ä+h #6[f:]+ +was weiss Ich=keine Ahnung;+

⁷ Die multimodale Transkription folgt Mondada (2019).

+dreht kopf nach re+

+dreht kopf nach li

+mehrmaliges kopfschütteln



#5



#6

+venezuela mexiko+ +is mir jetzt au WURSCHT;+

+bewegt mikro zwei mal von re nach li

+kleines schnelles kopfschütteln+

257

#7°h wie kann das sein [...]



#7

Nachdem er seinen Namen genannt und sich vorgestellt hat (im Transkript nicht reproduziert), kündigt der Sprecher eine Frage an (Z. 254). Bevor er allerdings die Frage stellt (Z. 260ff.) äußert er eine Behauptung (Z. 255), die er anschließend durch Beispiele spezifiziert (Z. 256-258). Diese Spezifizierung erfolgt jedoch nicht geradlinig: nach einer Mikropause beginnt die Konstruktion mit „so“, das hier unterschiedliche Funktionen erfüllen könnte. Die Formulierung wird dann unterbrochen durch eine Zögerungspartikel, den gelängten Frikativ [f:] und zwei epistemische Disclaimer. Die beiden nachfolgenden Beispiele (Venezuela und Mexiko) werden somit als approximativ und unsicher markiert. Die explizite Positionierung des Sprechers „is mir jetzt au wurscht“ zeigt, dass es sich nicht nur um Ungenauigkeit bzw. Unsicherheit handelt, sondern auch, wie sich der Sprecher zu dieser Ungenauigkeit positioniert.

Interessant ist in diesem Zusammenhang ein Blick auf die körperliche Realisierung der Formulierungssuche und der Positionierung. Während Kopfhaltung und Blickrichtung des Sprechers zunächst sehr wahrscheinlich auf die Experten auf dem Podium gerichtet sind

(#4), beginnt mit dem „so“ eine Kopfbewegung und eine Änderung der Blickrichtung, zunächst nach rechts (#5) und anschließend nach links (#6). Dieses relativ langsame Drehen des Kopfes geht über in ein Kopfschütteln und in eine Hin- und Her-Bewegung des Mikrofons während der Nennung der beiden Beispiele (Venezuela und Mexiko). Die Positionierung („is mir jetzt au wurscht“) wird von einem kleinen, sehr schnellen Kopfschütteln begleitet. Noch vor dem hörbaren Einatmen und der Frage befinden sich Kopf und Blick wieder in einer stabileren Haltung, nach vorn und auf das Podium gerichtet. Diese, zeitlich genau auf den verbalen Turn abgestimmten Bewegungen verdeutlichen nicht nur den Sonderstatus der Parenthese in Z. 257-259, sondern auch die Positionierung im Sinne eines Disengagement, genauer gesagt, einer Markierung von Desinteresse bezüglich der Genauigkeit in der Formulierung (siehe auch Baldauf-Quilliatre 2019), angekündigt durch [f:].

Während also der Sprecher im vorangegangenen Beispiel 3 ([pf:] mit schwachem Verschluss) kein weiteres Indiz gibt, dass er nicht wirklich in die Suche nach dem bestmöglichen Element engagiert ist, handelt es sich in Beispiel 4 ([f:]) um ein relativ eindeutiges Anzeichen von Desinteresse. Interessant ist hier nun, dass in beiden Fällen entweder nur ein schwacher oder gar kein Verschluss vorhanden ist. Gleichzeitig wird in den gesamten 553 untersuchten Okkurrenzen in keiner Sequenz dieser Art (Anzeigen von Formulierungsproblemen mit oder ohne Disengagement) ein Verschlusslaut (Plosiv oder Trill) ohne Frikativ benutzt.

Schwieriger sind Aussagen hinsichtlich der Länge des Frikativs in der Variante ohne Verschluss. Produziert ein Sprecher oder eine Sprecherin nur einen frikativischen Laut ist dieser in der Regel gelängt oder kurz mit relativ starker Intensität. Länge und Intensität des Frikativs variieren auch, wenn es sich um die Lautverbindung mit Plosiv oder Trill handelt. In beiden Fällen kann der Laut mit unterschiedlicher Intensität und in unterschiedlicher Länge realisiert werden, wobei labiodentale Frikative wohl nur dann kurz sind, wenn sie mit höherer Intensität produziert werden. Bilabiale Frikative weisen dagegen nur selten eine hohe Intensität auf, unabhängig davon, ob sie lang oder kurz sind. Hinsichtlich der sequenziellen Einbettung und Verwendung lassen sich, vergleichbar zu den Untersuchungen des Plosivs, nur Tendenzen aufzeigen. So sind etwa die Frikative in der Regel (aber nicht systematisch) relativ stark gelängt, wenn es sich um eine Positionierung des Sprechers handelt. Wird das Lautobjekt als Marker für den Abschluss einer Sequenz benutzt, ist der Frikativ in der Regel sehr kurz und schwach. Ausschnitt 5 zeigt ein solches Beispiel und kann bis zu einem gewissen Grad mit dem ersten Beispiel (entstimmlicher Trill) verglichen

werden. Es handelt sich um einen Ausschnitt aus einem Coaching, in dem TB, ein ehemaliger Pharmareferent eine neue Anstellung sucht und dafür die Hilfe eines Coachs (TN)⁸ in Anspruch nimmt.

Beispiel 5 FOLK_E_00173_SE_01_T_05

[https://dgd.ids-](https://dgd.ids-mannheim.de/DGD2Web/jsp/DisplayVideoTranscriptExcerpt.jsp?v_session_id=11C016FFD1DB3268E42D97462BEA0DF8&v_doc_id=FOLK_E_00173_SE_01_T_05_DF_01&v_contribution_id=c175&v_word_id=w896&v_context_ids=&v_context_size=4&v_text_size=200&v_download=true)

[mannheim.de/DGD2Web/jsp/DisplayVideoTranscriptExcerpt.jsp?v_session_id=11C016FFD1DB3268E42D97462BEA0DF8&v_doc_id=FOLK_E_00173_SE_01_T_05_DF_01&v_contribution_id=c175&v_word_id=w896&v_context_ids=&v_context_size=4&v_text_size=200&v_download=true](https://dgd.ids-mannheim.de/DGD2Web/jsp/DisplayVideoTranscriptExcerpt.jsp?v_session_id=11C016FFD1DB3268E42D97462BEA0DF8&v_doc_id=FOLK_E_00173_SE_01_T_05_DF_01&v_contribution_id=c175&v_word_id=w896&v_context_ids=&v_context_size=4&v_text_size=200&v_download=true)

- | | | |
|-----|----|---|
| 178 | TN | [...] vor (.) DEM hintergrund würd ich's glaub ich
[dann eher so machen.] |
| 180 | TB | [((schmatzt)) genAUso seh ich_s AU]CH.=ja. |
| 181 | | (0.60) |
| 182 | TN | dann hab ich zwar wieder ne KÜNdigungsfrist aber [pf:] |
| 183 | TB | °h |
| 184 | | (0.20) |
| 185 | TB | JA was aber [entschEI]dend is, [...] |

Der Ausschnitt beginnt mit dem Ende einer Sequenz, in der sich die beiden Partner daraufhin einigen, dass es besser ist, zunächst einmal eine Stelle anzunehmen, auch wenn sie nicht vollkommen den eigenen Wünschen und Ansprüchen entspricht. Nach einer Pause öffnet nun TN in Z. 182 eine neue Sequenz mit einem Einwand – der Konnektor „aber“ projiziert hier eine adversative Konstruktion, in welcher der Einwand aufgehoben wird. Diese Aufhebung wird jedoch nicht expliziert, TN beendet seinen Turn mit einem kurzen, 0,18 Sekunden dauernden [pf], das sich phonetisch durch eine relativ geringe Intensität und ein hörbares Nachlassen der Spannung auszeichnet. Die Lautverbindung befindet sich nicht im Overlap, ein Sprecherwechsel wird jedoch gleich anschließend mit TB's hörbarem Einatmen angekündigt und TB formuliert einen neuen Einwand – die Kündigungsfrist ist somit nicht mehr aktuell. Außerdem befindet sich [pf] hier an der Stelle einer Proposition, die bereits expliziert wurde (*es ist besser, sich nicht aus der Arbeitslosigkeit heraus zu bewerben*). Insofern handelt es sich weder um eine *trail-off* Konjunktion (Walker 2012), noch um einen Turnabbruch und eine Einladung an die anderen Teilnehmer*innen, nach weiteren, nicht explizierten Inhalten zu suchen (Raymond 2004 für englisches „so“ in dieser Position). Im

⁸ Für eine genauere Analyse siehe Baldauf-Quilliatre / Imo (2020).

Gegenteil, [pf] lädt vielmehr dazu ein, in den vorangegangenen Turns nach dem nicht explizierten Argument zu suchen.

Gemeinsam hat dieses Beispiel mit dem ersten Ausschnitt aus phonetischer Sicht die Fokussierung auf den Verschlusslaut, wobei es sich in Beispiel 1 um einen mit relativ hoher Intensität produzierten Trill handelt (deutlich sichtbar auch in den aufgeblähten Wangen und gestülpten Lippen), während in Beispiel 5 auch der bilabiale Plosiv [p] eher schwach und kurz ist. Sequenziell betrachtet, handelt es sich in Beispiel 1 um die Ankündigung einer Positionierung innerhalb eines von längeren Pausen unterbrochenen Turns, während Beispiel 5 den Abschluss eines Turns markiert und auf vorangegangene Sequenzen rückverweist.

Ein letzter Ausschnitt aus einem Interview zeigt einen langen bilabialen Frikativ in einer typischen Verwendungsweise. AH erzählt von seiner Zeit als Malerlehrling und erwähnt dabei ein Gespräch, indem er angegeben hatte, dass er gern etwas anderes gelernt hätte.

Beispiel 6 FOLK_E_00325_SE_01_T_01

https://dgd.ids-mannheim.de/DGD2Web/jsp/DisplayTranscriptExcerpt.jsp?v_session_id=11C016FFD1DB3268E42D97462BEA0DF8&v_doc_id=FOLK_E_00325_SE_01_T_01_DF_01&v_contribution_id=c1014&v_word_id=w3294&v_context_ids=&v_context_size=4&v_text_size=200&v_download=true

[...]
1011 AH sei doch froh dAss schon mal ein beruf hast andere ham GAR kein
 beruf und- [°hh]
1012 ML [hmhm]
1013 (0.36)
1014 na ja un[d [pϕ:] hm gr]oß vertEIdigen konnt ich mich NICH,
1015 [um zu zeigen]
1016 AH hmhm;
1017 (0.42)

Nachdem AH in Zeile 1011 eine Redewiedergabe inszeniert, erklärt er in Zeile 1014-1016 sein damaliges Verhalten aus gegenwärtiger Sicht. Die Erklärung wird eingeleitet durch die Lautverbindung [pϕ:], bestehend aus einem relativ schwachen Verschluss und einem sehr langen Frikativ mit abnehmender Intensität, mit einer Dauer von insgesamt ca. 0,87 Sekunden. Im Unterschied zu Beispiel 3, wo die Lautverbindung ebenfalls turnmedian auftritt, gibt es hier kein Anzeichen von Formulierungsschwierigkeiten: der Turn weist weder

Wiederholungen oder Reformulierungen, noch Pausen oder andere Zögerungssignale auf. Auch findet man keine lexikalischen Anzeichen für Vagheit oder Hinweise auf eine gemeinsame Suche nach dem am besten geeigneten Element. In Hinblick auf andere hier diskutierte oder bereits beschriebene Verwendungen, u.a. als Einleitung einer Bewertung, lässt sich [pɸ:] in diesem Kontext als Anzeichen für eine Positionierung beschreiben. Da keine der anderen Teilnehmerinnen hörbar darauf reagiert, ist die Art der Positionierung nicht zu bestimmen, allerdings weist die nachfolgende Proposition („groß verteidigen konnt ich mich nich“) und der in anderen sequenziellen Kontexten beschriebene Zusammenhang zu Desinteresse und Disengagement auf eine Art Defätismus hin.

Fazit

In Hinblick auf die zu Anfang des Abschnitts gestellten Frage nach (1) der Existenz eines Lautobjekts „pff“ mit eventuell verschiedenen Funktionen oder mehrerer verschiedener Lautobjekte und (2) der Funktion einer phonetischen Transkription im Rahmen der Gesprächsanalyse lässt sich Folgendes feststellen:

Betrachtet man nur die in FOLK als „pff“ transkribierten Okkurrenzen (entsprechend der eingangs gezeigten Kategorisierung, siehe Schmidt et al. 2015 und Winterscheidt et al. 2019), zeigt eine genauere phonetische Transkription, dass hier eine Reihe verschiedener Lautverbindungen aus Plosiv/Trill und Frikativ zusammengefasst werden, wobei es sich zum Teil sogar nur um einen einzelnen Laut (Plosiv/Trill oder Frikativ) handelt. Beide Segmente weisen Unterschiede hinsichtlich Intensität und Aspiration (Plosiv/Trill) bzw. Intensität und Länge (Frikativ) auf.⁹ Eine Sequenzanalyse verdeutlicht, dass es bestimmte Tendenzen bezüglich der Verwendung bestimmter Varianten in bestimmten Kontexten gibt, aber eine deutliche kategoriale Trennung scheint nicht zu bestehen (allerdings müsste diese Aussage durch systematischere interaktionsphonetische Analysen bestätigt werden). Das lässt nun den Schluss zu, dass es eine Gruppe von stimmlosen konsonantischen egressiven vorderen (labialen) Lautverbindungen gibt, die in bestimmten sequenziellen Kontexten auftreten. Insofern ist die Annahme einer Lautobjekts „pff“ durchaus zutreffend, allerdings kann nur die Berücksichtigung der phonetischen Unterschiede diese Annahme auf ein solides

⁹ Aus Platzgründen wurden hier nur zwei Aspekte näher betrachtet: die Intensität (der Verschlusslaute) und die Länge (der Frikative). Weitere Kombinationen wie etwa Intensität und Aspiration in Bezug auf Verschlusslaute oder Intensität bzw. Intensität und Länge in Bezug auf Frikative wurden nicht im Detail untersucht.

Fundament stellen – inwiefern „pft“ (mit abschließendem Verschlusslaut) tatsächlich eine eigene Form darstellt oder ebenfalls zu den Varianten von „pff“ gehört (und dann auch keiner besonderen Transkription bedarf), ist hier nicht untersucht wurden.

Die Annahme einer Gruppe von Lautverbindungen, die sowohl phonetische Ähnlichkeiten aufweisen, als auch in vergleichbaren sequenziellen Kontexten auftreten, wird durch andere Studien zu Lautobjekten bzw. Vokalisierungen in der Interaktion gestützt. So argumentiert etwa Ward (2006: 9), der versucht Laut-Bedeutungs-Korrespondenzen für englische, nicht-lexikalisierte Vokalisierungen aufzuzeigen, dass es sich dabei um nicht-kategoriale graduelle Formen handelt. Gleichzeitig wird darauf hingewiesen, dass die Variabilität und somit die Anpassungsfähigkeit an die lokale Umgebung zu den Charakteristika von nicht-lexikalisierten Vokalisierungen zählt (Keevallik & Ogden 2020: 5).

Während die phonetische Transkription also einerseits dazu beiträgt, die Annahme eines Lautobjekts „pff“ zu bestätigen, erlaubt sie andererseits eine Präzisierung hinsichtlich verschiedener Funktionen. So scheint ein Verschlusslaut notwendig, wenn es sich um das Anzeichen einer Positionierung handelt, er ist aber nicht notwendig, wenn Formulierungsschwierigkeiten angezeigt werden. Angaben dieser Art sind nun nicht nur für eine genauere Ausdifferenzierung eines Lautobjekts interessant, sondern auch in Bezug auf die Körperlichkeit des Sprechens. Zum einen kann die sichtbare Realisierung des Lautobjekts deren Wahrnehmung und Interpretation in der Interaktion beeinflussen. Ein sichtbarer Verschluss mit vorgestülpten Lippen und geblähten Wangen (Bsp. 1) ist für die anderen Interaktionsteilnehmer deutlich sichtbar und weist somit auch auf seine Bedeutung und seine Verantwortbarkeit (*accountability*) in der Interaktion hin. Zum anderen kann eine phonetisch exaktere Beschreibung sowohl die teilweise enge Verbindung zwischen Lautobjekt und physiologischen Prozessen, als auch deren interaktionale Nutzung genauer aufzeigen.

5. Schluss

Das gesprächsanalytische Transkriptionssystem GAT2 (Selting et al. 2008) sieht drei Granularitätsstufen vor (Minimaltranskript, Basistranskript und Feintranskript), die sich überwiegend hinsichtlich der Einbeziehung prosodischer Parameter unterscheiden. Das Basistranskript mit Angaben zu Fokusakzent und Tonhöhenbewegungen am Segmentende

gilt dabei als Standard für interaktionslinguistische Arbeiten. Vor allem interaktionsphonetische Arbeiten haben immer wieder auf die Bedeutung prosodischer Parameter für die Interaktionskonstitution hingewiesen und ihre grundsätzliche Berücksichtigung in der Transkription (und somit auch mehr oder weniger in einer Analyse, die nicht primär auf prosodische Phänomene zielt) resultiert daraus. Lautobjekte hingegen stehen erst seit kurzem im Fokus interaktionslinguistischer und besonders interaktionsphonetischer Arbeiten, auch wenn ihre Bedeutung für die Interaktionskonstitution bereits seit langem bekannt ist. Dies spiegelt sich auch in der Transkription wieder, da es hierzu kaum Angaben in Bezug auf „gute Forschungspraktiken“ gibt. Die hier vorgestellten Überlegungen sollen erste Schritte in diese Richtung sein.

Lautobjekte stellen die gesprächsanalytische Transkription vor mehrere Probleme. Usuelle und stärker lexikalisierte Formen besitzen in der Regel eine Schreibweise, die eine orthographische Transkription ermöglicht. Allerdings haben gesprächsanalytische Arbeiten gezeigt, dass insbesondere eine Interpretation ohne Bezug auf prosodische Parameter nicht möglich ist (siehe etwas „oh“ oder „hm“). Eine Notation ohne tonale Struktur kann daher meines Erachtens nur bei einer sehr groben Granularität möglich sein und erlaubt nur sehr begrenzte Aussagen zur Funktion des Lautobjekts in der Interaktion. Während für Tonstrukturen in GAT2 durchaus Lösungen angeboten werden (wenn auch nicht spezifisch für die Transkription von Lautobjekten), sind segmentalphonetische Parameter überhaupt nicht erwähnt. Dabei kann es durchaus wesentlich sein, wie ein Laut realisiert wird. Das trifft nun noch stärker auf weniger oder nicht-lexikalisierte Formen ohne konventionalisierte Schreibweisen zu. Welche phonetischen Parameter hier obligatorisch sind und welche nicht, ist oft (noch) nicht erwiesen. Bei stimmlosen Lautobjekten wie dem hier als Beispiel erwähnten „pff“, kommt der lautlichen Realisierung besonderes Gewicht zu, da bestimmte prosodische Aspekte wie Tonhöhenbewegungen nicht zum Tragen kommen. Eine relativ exakte Transkription der produzierten Laute ist daher zumindest in einem ersten Schritt notwendig. Ist der phonetische Bereich, in dem sich ein Lautobjekt bewegt, bekannt, spricht meines Erachtens nichts für die Einigung auf eine Schreibweise im Minimaltranskript.¹⁰ Basis- oder Feintranskripte allerdings sollten der Komplexität von Lautobjekten Rechnung

¹⁰ Ein Minimaltranskript erfasst Wortlaut und segmentale Unterteilung einschließlich Überlappungen, Verzögerungen, Pausen, Ein- und Ausatmen, Lachen, Nonverbales sowie Schwer- oder Unverständliches (Selting et al. 2019 : 359).

tragen und daher den Laut/die Lautverbindung so genau wie möglich beschreiben, und zwar sowohl in Bezug auf die prosodische als auch auf die segmentale Realisierung. Wenn das Internationale Phonetische Alphabet dafür entsprechende Zeichen zur Verfügung hat, ist es sinnvoll, diese auch zu nutzen.

Literatur

- Ayaß, Ruth. 2015. Doing data: The status of transcripts in Conversation Analysis. *Discourse Studies*, 17(5), 505–528.
- Baldauf-Quilliatre, Heike. 2018. Über Lautäußerungen und ihr Verhältnis zur Sprache (am Beispiel von *pff*). In Laurent Gautier et al. (Hrsg.): *Diskursive Verfestigungen. Schnittstellen zwischen Morphosyntax, Phraseologie und Pragmatik im Deutschen und im Sprachvergleich*, Berlin: de Gruyter, 293–310.
- Baldauf-Quilliatre, Heike. 2019. *Entre langage et extralinguistique: Pour une intégration des phénomènes vocaux non-lexicalisés dans une conception du langage – l'exemple de „pff“ en allemand et en français*, Unveröffentlichte Habilitationsschrift.
- Baldauf-Quilliatre, Heike & Imo, Wolfgang. 2020. *pff*. In: Wolfgang Imo & Jens Lanwer (Hrsg.): *Prosodie und Konstruktionsgrammatik*. Berlin: de Gruyter, 201–232.
- Baldauf-Quilliatre, Heike et al. 2014. oh là là - The contribution of the multimodal database CLAPI. In Henry Tyne et al. (Hrsg.): *French through Corpora: Ecological and Data-Driven Perspectives in French Language Studies*. Newcastle: Cambridge Publishers, 167–197.
- Balnat, Vincent. 2012. was is der hässliche zwerch?? ^^g – Anpassungsstrategien im Chat. In Anne Larrory-Wunder (Hrsg.): *Intersubjektivität und Sprache. Zur An- und Abgleichung von Sprecher- und Hörervorstellungen in Texten und Gesprächen*, Tübingen: Stauffenburg, 91-116.
- Barth-Weingarten, Dagmar. 2011. Double sayings of German JA—More observations on their phonetic form and alignment function. *Research on language and social interaction*, 44(2), 157–185.
- Bolden, Galina. 2015. Transcribing as research: “manual” transcription and conversation analysis. *Research on Language and Social Interaction*, 48(3), 276–280.
- Bucholtz, Mary. 2007. Variation in transcription. *Discourse studies*, 9(6), 784–808.
- Davidson, Christina. 2010. Transcription matters: Transcribing talk and interaction to facilitate conversation analysis of the taken-for-granted in young children's interactions. *Journal of Early Childhood Research*, 8(2), 115–131.
- Deppermann, Arnulf. 2001. *Gespräche analysieren*. Opladen: Leske+ Budrich.
- Deppermann, Arnulf & Schütte, Wilfried. 2008. Data and transcription. In Gerd Antos & Elja Ventola (Hrsg.): *Handbook of interpersonal communication*, Berlin: de Gruyter, 179–213.
- Dingemanse, Mark. 2020. Between Sound and Speech: Liminal Signs in Interaction. *Research on Language and Social Interaction*, 53(1), 188–196.
- Ehlich, Konrad. 2007. Interjektion und Responsiv. In Ludger Hoffman (Hrsg.): *Deutsche Wortarten*. Berlin: de Gruyter, 423–444.
- Fauré, Laurent, & Verine, Bertrand. 2001. Authentifier un discours autre en y mettant du sien: les vocalisations ah et oh en frontière de discours rapporté direct à l'oral. In Juan-

- Manuel Lopez-Munoz et al. (Hrsg.): *Le discours rapporté dans tous ses états*. Paris: L'Harmattan, 317–327.
- Gardner, Rod. 2001. *When Listeners Talk*, Amsterdam, Philadelphia: Benjamins.
- Goffman, Erving. 1981. *Forms of talk*. University of Pennsylvania Press.
- Golato, Andrea & Fagyal, Zsuzsanna. 2008. Comparing single and double sayings of the German response token ja and the role of prosody: A conversation analytic perspective. *Research on language and social interaction*, 41(3), 241–270.
- Goodwin, Charles. 1986. Between and within: Alternative sequential treatments of continuers and assessments. *Human Studies* 9, 205–217.
- Gonzalez Temer, Veronica (2017). *A multimodal analysis of assessment sequences in Chilean Spanish interaction*. PhD Thesis, University of York.
<http://etheses.whiterose.ac.uk/20579/> (zuletzt aufgerufen am 20.07.2021)
- Gülich, Elisabeth & Kotschi, Thomas. 1996. Textherstellungsverfahren in mündlicher Kommunikation. Ein Beitrag am Beispiel des Französischen. In Wolfgang Motsch (Hrsg.) : Ebenen der Textstruktur. Sprachliche und kommunikative Prinzipien. Tübingen: Niemeyer, 37–80.
- Hepburn, Alexa. 2004. Crying: Notes on description, transcription, and interaction. *Research on Language and Social Interaction*, 37(3), 251–290.
- Hepburn, Alexa & Bolden, Galina (2017). *Transcribing for social research*. London: Sage.
- Hoey, Elliott. 2020. Waiting to inhale: On sniffing in conversation. *Research on Language and Social Interaction*, 53(1), 118–139.
- Hofstetter, Emily. 2020. Non-lexical ‘moans’: Response cries in board game interactions. *Research on Language and Social Interaction*, 53(1), 42–65.
- ICOR. 2008. “Oh::, oh là là, oh ben...”, les usages du marqueur “oh” en français parlé en interaction. *Actes du Congrès Mondial de Linguistique Française (CMLF08)*.
<https://doi.org/10.1051/cmlf08099> (zuletzt aufgerufen am 20.07.2021)
- Jefferson, Gail. 1985. An exercise in the transcription and analysis of laughter. In Teun van Dijk (Hrsg.): *Handbook of discourse analysis, Vol. 3: Discourse and dialogue*, London: Academic Press, 25–34.
- Jefferson, Gail. 2004. Glossary of transcript symbols In Gene Lerner (Hrsg.): *Conversation analysis: Studies from the first generation*. Amsterdam: Benjamins, 13–31.
- Keating, Patricia 1988. Underspecification in phonetics. *Phonology*, 5(2), 275–292.
 10.1017/S095267570000230X (zuletzt aufgerufen am 20.07.2021)
- Mondada, Lorenza. 2007. Commentary: Transcript variations and the indexicality of transcribing practices. *Discourse studies*, 9(6), 809–821.
- Mondada, Lorenza. 2018. Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality. *Research on Language and Social Interaction*, 51(1), 85–106.
- Mondada, Lorenza. 2019. Conventions for multimodal transcription (erste Version: 2001; Aktuelle Version: 5.0.1, Nov. 2019), <https://www.lorenzamondada.net/multimodal-transcription> (zuletzt aufgerufen am 20.07.2021).
- Mondada, Lorenza. 2020. Audible sniffs: Smelling-in-interaction. *Research on Language and Social Interaction*, 53(1), 140–163.
- Nübling, Damaris. 2004. Die prototypische Interjektion: Ein Definitionsvorschlag. *Zeitschrift für Semiotik*, 26(1-2), 11–46.
- Keevallik, Leelo & Ogden, Richard. 2020. Sounds on the Margins of Language at the Heart of Interaction. *Research on Language and Social Interaction*, 53(1), 1–18.

- Pompino-Marschall, Bernd. 2004. Zwischen Tierlaut und sprachlicher Artikulation: Zur Phonetik der Interjektionen. *Zeitschrift für Semiotik*, 26(1-2), 71–84.
- Potter, John & Hepburn, Alexa. 2010. Putting aspiration into words: 'Laugh particles', managing descriptive trouble and modulating action. *Journal of Pragmatics*, 42(6), 1543–1555.
- Proske, Nadine. 2014. „h ach KOMM; hör AUF mit dem kIEInkram“. Die Partikel *komm* zwischen Interjektion und Diskursmarker. *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion*, 15, 121–160 www.gespraechsforschung-ozs.de (zuletzt aufgerufen am 20.07.2021).
- Psathas, George & Anderson, Timothy 1990. The ‚practices‘ of transcription in conversation analysis. *Semiotica*, 78(1-2), 75–100.
- Raymond, Geoffrey. 2003. Grammar and social organization: Yes/no interrogatives and the structure of responding. *American sociological review*, 939–967.
- Raymond, Geoffrey. 2004. Prompting action: The stand-alone "so" in ordinary conversation. *Research on Language and Social Interaction*, 37(2), 185–218.
- Reber, Elisabeth. 2012. Affectivity in interaction. *Sound objects in English*. Amsterdam: Benjamins.
- Reber, Elisabeth & Couper-Kuhlen, Elizabeth. 2010. Interjektionen zwischen Lexikon und Vokalität: Lexem oder Lautobjekt. In Arnulf Deppermann & Angelika Linke (Hrsg.): *Sprache intermedial: Stimme und Schrift, Bild und Ton*, Berlin: de Gruyter, 69–96.
- Reber, Elisabeth & Couper-Kuhlen, Elizabeth. 2020. On 'whistle' sound objects in English everyday conversation. *Research on Language and Social Interaction*, 53(1), 164–187.
- Schegloff, Emmanuel. 1982. Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In Deborah Tannen (Hrsg.): *Analyzing discourse: Text and talk*, Washington: Georgetown University Press, 71–93.
- Schmidt, Thomas et al. 2015. *cGAT. Konventionen für das computergestützte Transkribieren in Anlehnung an das Gesprächsanalytische Transkriptionssystem 2 (GAT2)*. https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/4616/file/Schmidt_Schuette_Winterscheid_cGAT_2015.pdf (zuletzt aufgerufen am 20.07.2021)
- Selting, Margret et al. 2009. Gesprächsanalytisches Transkriptionssystem (GAT2). *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion*, 10, 353–402 www.gespraechsforschung-ozs.de (zuletzt aufgerufen am 20.07.2021)
- Szymanski, Margaret. 1999. Re-engaging and dis-engaging talk in activity. *Language in Society*, 28, 1–23.
- Ten Have, Paul. 2002. Reflections on transcription. *Cahiers de praxématique*, 39, 21–43.
- Wiggins, Sally. 2002. Talking with your mouth full: Gustatory mmms and the embodiment of pleasure. *Research on language and social interaction*, 35(3), 311–336.
- Wiggins, Sally & Keevallik, Leelo. 2021. Enacting Gustatory Pleasure on Behalf of Another: The Multimodal Coordination of Infant Tasting Practices. *Symbolic Interaction*, 44(1), 87–111.
- Winterscheid, Jenny et al. (2019) *Normalisieren mit OrthoNormal. Konventionen und Bedienungshinweise für die orthografische Normalisierung von FOLKER-Transkripten*. https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/9326/file/Winterscheid_etal_Normalisierungskonventionen_2019.pdf (zuletzt aufgerufen am 20.07.2021)

- Walker, Gareth. 2012. Coordination and interpretation of vocal and visible resources: 'Trail-off' conjunctions. *Language and speech*, 55(1), 141–163.
- Ward, Nigel. 2006. Non-lexical conversational sounds in American English. *Pragmatics & Cognition*, 14(1), 129–182.

Dr. habil Heike Baldauf-Quilliatre

Forschungsschwerpunkte: multimodale Konversationsanalyse / Interaktionale Linguistik, Vokalität, Lautobjekte, Spielinteraktion, Positionierung, Teilnehmerkategorien

Universität Lyon 2 Lumière, Faculté des Langues

Forschungszentrum ICAR, ENS de Lyon, 15 parvis René Descartes, BP 7000, 69342 Lyon Cedex 7, Frankreich

heike.baldauf@univ-lyon2.fr

<http://www.icar.cnrs.fr/membre/hbaldauf-quilliatre/>