

Medienlinguistische Methodik

Datenerhebung durch Beobachtung und Befragung

Arne Rubehn

Lehrstuhl für Multilinguale Computerlinguistik
Universität Passau

28.10.2025



Überblick

In dieser Sitzung widmen wir uns der **Erhebung von wissenschaftlichen Daten** durch **Beobachtung oder Befragung**.

Welche grundlegenden Verfahren der Datenerhebung gibt es?

Wie funktionieren diese Verfahren im Wesentlichen?

Was sind die jeweiligen Vor- und Nachteile?



Wiederholung

Daten liefern **Informationen** zu Objekten und Prozessen in der Wissenschaft. Im engeren Sinne sollten sie **maschinenlesbar** und (relativ) einheitlich **strukturiert** sein.

Wir stellen den Anspruch, dass wissenschaftliche Vorhaben **ehrlich, offen** und **falsifizierbar** sind.

Die Erhebung und Analyse von wissenschaftlichen Daten sollte daher **objektiv, verlässlich** und **repräsentativ** sein.

Was bedeutet das für euch?



Wiederholung

Die Erhebung und Analyse von wissenschaftlichen Daten sollte **objektiv, verlässlich** und **repräsentativ** sein.

Objektivität. Unabhängigkeit & Wertfreiheit von der durchführenden Person.

Reliabilität. Konstantes Ergebnis bei wiederholter Messung bzw. Erhebung.

Repräsentativität. Daten bilden zugrundeliegende Objekte oder Prozesse möglichst realistisch ab.



Verfahren der Datenerhebung

Beobachtung. Teilnehmende Beobachtung und nachfolgende Protokollauswertung.

Befragung. Systematische Befragung anhand von Leitfäden oder Fragebögen.

Experiment. Messung bestimmter Parameter unter stark kontrollierten Bedingungen (z.B. Eyetracking, EEG, MRT, ...)

Korpussurvey. Zusammenstellung von (Text-)Korpora für qualitative oder quantitative Analysen.



Verfahren der Datenerhebung

Beobachtung. Teilnehmende Beobachtung und nachfolgende Protokollauswertung.

Befragung. Systematische Befragung anhand von Leitfäden oder Fragebögen.

Experiment. Messung bestimmter Parameter unter stark kontrollierten Bedingungen (z.B. Eyetracking, EEG, MRT, ...) *[nächste Woche]*

Korpussurvey. Zusammenstellung von (Text-)Korpora für qualitative oder quantitative Analysen. *[übernächste Woche]*



Metasprache und Objektsprache

Bevor wir uns den verschiedenen Erhebungsmethoden widmen, ist zunächst einmal eine grundlegende Unterscheidung zwischen zwei Arten von Sprachdaten zu treffen.

Metasprache. Sprechen über Sprache.

z.B. *Wie häufig sprichst du Dialekt?*

Objektsprache. Sprache als Forschungsobjekt.

z.B. *Wie nennst du ein Brötchen in deinem Dialekt?*



Beobachtung & Befragung

Beobachtungen und **Befragungen** sind geeignete Mittel, direkt *primäre* Sprachdaten zu generieren. Für Experimente oder Korpus surveys ist typischerweise eine Basis an Primärdaten notwendig.

Gerade in der Forschung zu “kleinen” Sprachvarietäten (Minderheitensprachen, Dialekte, Soziolekte) spielen Beobachtungen und Befragungen daher eine große Rolle. Sie sind für die **Sprachdokumentation** und die **Variationslinguistik** unabdingbar.



Beobachtung



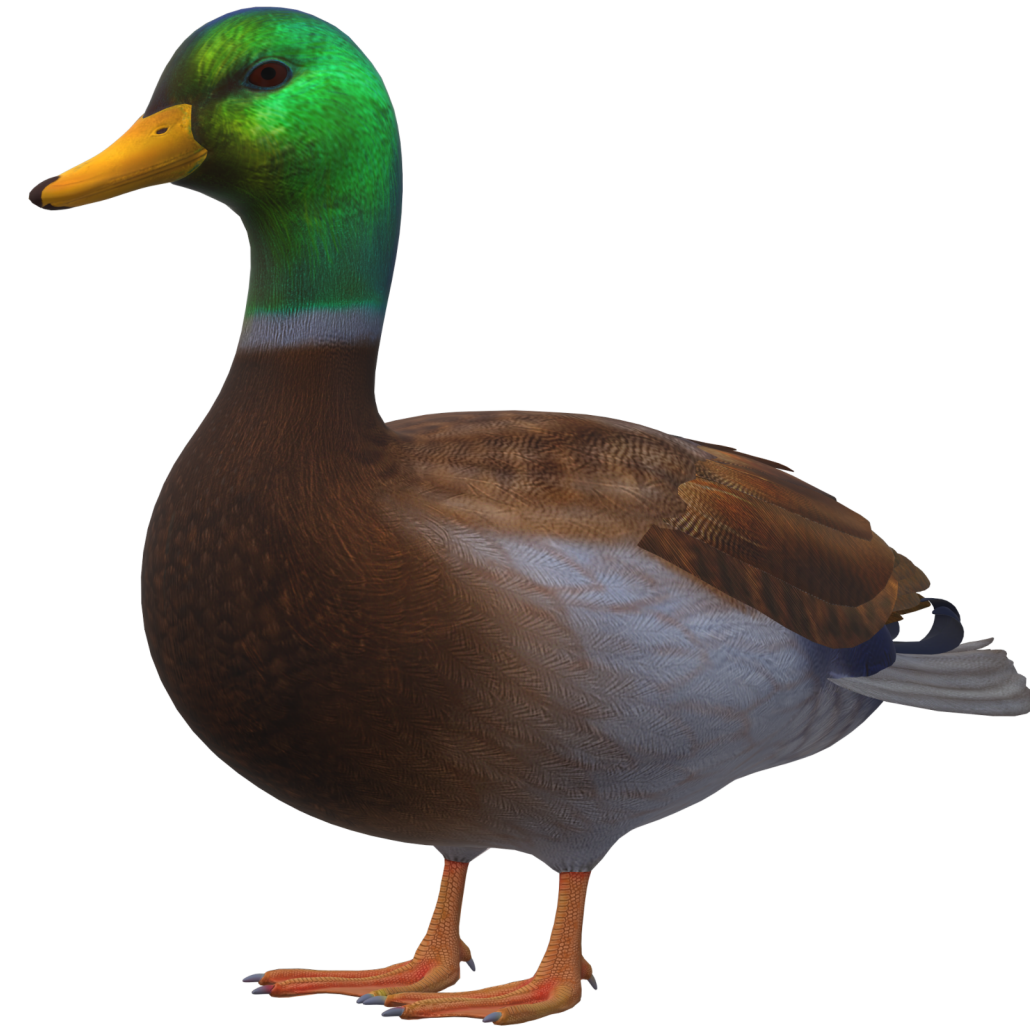
Untersuchungsgebiet : Stadtwaldsee Bremen 04.06.2025,
Repräsentatives, urbanes Binnengewässer mit unterschiedlichen Ufervegetationstypen

Vegetationstypen :

- **Vege1** : Offenes Ufer - kaum Vegetation, Charakterisiert durch geringe Deckung, fehlende natürliche Schutzstrukturen, Anfällig für Degradation durch Ufernutzung^[1, S. 81, S. 94]
- **Vege2** : Schilfblattzone - mittlere Vegetation, Repräsentiert strukturreiche Uferbereiche (Röhrichtgürtel)^[1, S. 139, S. 144; 2, 667; 3, S. 347]
- **Vege3** : Ufergehölzstreifen - Repräsentiert strukturreiche Uferbereiche (Röhrichtgürtel)^[1, S. 139, S. 144; 2, 667; 3, S. 347].

Datenerhebung : 27 Datenaufnahmen für jede Vegetationszone, erfasst Anzahl und Art der Entenvögel, in 9 Felder aufgeteilt⁸

Einfluss von Vegetation auf die Artenvielfalt von Entenvögeln (Anatidae)



Beobachtung

Die wohl “simpelste” Form der Datenerhebung ist die **Beobachtung**. Manche Forschungsobjekte lassen sich mit bloßem Auge beobachten (z.B. Tiere oder Bäume), für andere sind technische Hilfsmittel erforderlich (z.B. Bakterien oder Planeten).

Für gute Forschungsdaten sind **systematische Beobachtungen** notwendig, die sich an der Forschungsfrage orientieren.

Im linguistischen Kontext findet Beobachtung typischerweise im Rahmen von **Feldforschung** statt: Forschende verbringen einige Zeit in einer Community und **dokumentieren** diese (Video-/Audioaufnahmen, Transkripte, Notizen, ...).



Fallstudie: Die “türkischen Powergirls”

Soziolinguistische Studie von Inken Keim mit dem Titel:

Deutsch-türkische Sprachvariation und die Herausbildung kommunikativer und sozialer Stile in jugentlichen MigrantInnengruppen türkischer Herkunft in Mannheim.

Keim untersucht eine Gruppe **deutsch- und türkischsprachiger junger Frauen** aus Mannheim, deren Eltern aus der Türkei nach Deutschland migriert sind, in der Übergangsphase vom Jugend- in das junge Erwachsenenalter.



Fallstudie: Die “türkischen Powergirls”

Als hauptsächliche Datengrundlage dienen Aufzeichnungen einer **teilnehmenden Beobachtung**, bei der Keim die jungen Frauen in ihrem Alltag begleitet.

Keim gewinnt das Vertrauen der Gruppe und kann dadurch **authentische Sprachdaten** aus Gesprächen der Frauen untereinander und mit anderen Personen (Familie, Lehrer:innen, ...) aufzeichnen.

Im Mittelpunkt ihrer Untersuchung steht eine qualitative Analyse dessen, wie die bilingualen Sprecherinnen verschiedene Stile als **soziales Zeichen** einsetzen.



Befragung



<https://partici.fi/92489810>



Befragung

Auch **Befragungen** sind ein probates Mittel, um Sprachdaten zu erheben. Hierbei können wir feiner zwischen Fragebögen und Interviews unterscheiden.

Fragebögen bestehen aus einer Reihe an vorgefertigten, typischerweise geschlossenen Fragen. Sie können von Gewährspersonen selbstständig und ohne Aufsicht der Forschenden ausgefüllt werden.

Interviews sind Gespräche zwischen Gewährspersonen und Forschenden. Sie sind üblicherweise lose strukturiert und orientieren sich an Leitfäden.



Befragung

Bei Befragungen müssen einige Fragen vorab geklärt werden.

Wer wird befragt? Gewährspersonen sollten in der Lage sein, **authentische** und **repräsentative Daten** zu liefern.

z.B.: Dialektdaten werden üblicherweise von *sesshaften* Sprecher:innen erhoben.

Was wird gefragt? Die Gestaltung des Fragebogens oder Interviewleitfadens richtet sich nach der **Forschungsfrage**; elizitierte Daten sollten dafür geeignet sein.

z.B.: Für die historische Linguistik ist vor Allem der Grundwortschatz relevant.



Befragung

Bei Befragungen müssen einige Fragen vorab geklärt werden.

Wie wird gefragt? Es gibt **direkte** und **indirekte Erhebungsmethoden**, z.B.:

direkte Befragung: *Wie sagst du zu "Tasse"?*

Benennung (indir.): *Wie nennt man den Nachwuchs von Hunden?* (Welpen).

Vervollständigung (indir.): *Man süßt seinen Tee mit ...* (Zucker)

Was könnten die Vorteile von indirekten Erhebungsmethoden sein?



Befragung

Bei Befragungen müssen einige Fragen vorab geklärt werden.

Wie wird gefragt? Es gibt **direkte** und **indirekte Erhebungsmethoden**, z.B.:

direkte Befragung: *Wie sagst du zu "Tasse"?*

Benennung (indir.): *Wie nennt man den Nachwuchs von Hunden?* (Welpen).

Vervollständigung (indir.): *Man süßt seinen Tee mit ...* (Zucker)

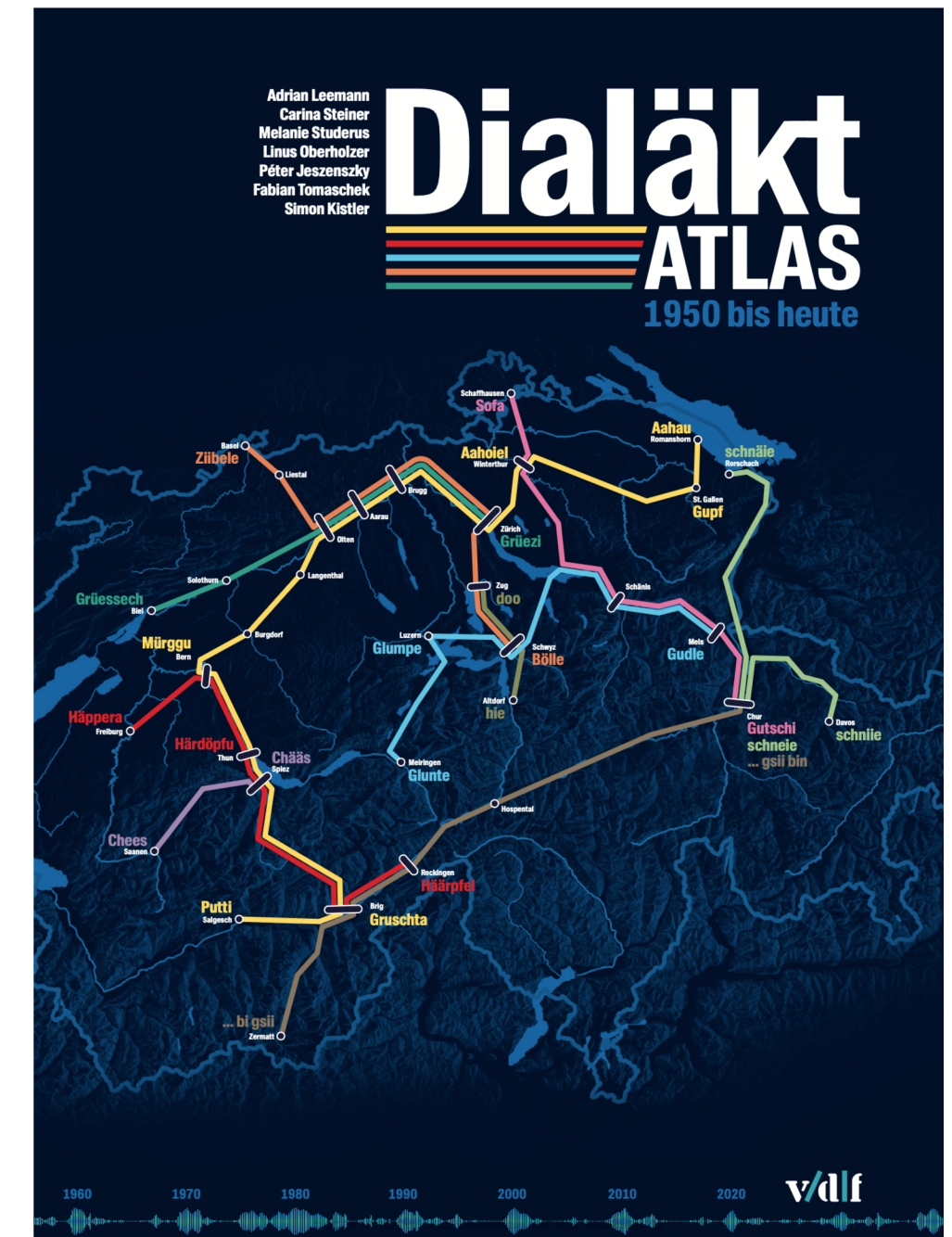
Indirekte Befragungsmethoden liefern häufig authentischere Daten, da sie **Priming-Effekte** vermeiden. Allerdings machen sie es auch schwerer, das gewünschte Objekt zu elizitieren.



Befragung

Lest § 3.1-3.4 in Leemann et al. (2024, S. 14-19). Welche Entscheidungen werden *vor* der Befragung getroffen und warum? Nach welchen Kriterien erfolgt die Auswahl der Gewährpersonen, der elizitierten Objekte und des Erhebungsverfahrens?

Trifft das beschriebene Erhebungsverfahren am Ehesten auf die Beschreibung der *Beobachtung*, des *Interviews* oder des *Fragebogens* zu?



Beobachtung & Befragung

Wir sehen, dass die Einteilung in Beobachtungen, Interviews und Fragebögen **keine Kategorisierung** mit klaren Grenzen ermöglicht. Da sich einzelne Elemente frei kombinieren lassen, kann man sich das Ganze eher als **Spektrum** vorstellen.

lose Struktur

feste Struktur



Beobachtung

Interview

Fragebogen

reichhaltige Daten
Spezifizierung von besonderen Phänomenen
Kategorisierungen *nach* der Analyse

vergleichbare Daten
Generalisierung von typischen Phänomenen
Kategorisierungen *vor* der Analyse



Beobachterparadoxon

Forschende sind an möglichst **authentischer Sprache** interessiert.

“To obtain the data most important for linguistic theory, we have to observe how people speak when they are not being observed.”

Daraus ergibt sich das sogenannte **Beobachterparadoxon**:

“The aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation.”



Fallstudie: Foxy Boston

Teilstudie im Rahmen einer **variationslinguistischen Erhebung** des Sprachgebrauchs einer multiethnischen Gruppe in East Palo Alto, Kalifornien.

Klassisches **soziolinguistisches Interview** (dessen Probleme kritisch diskutiert werden).

Untersuchung von zwei Interviews (1990 & 1991) mit **Foxy Boston** (Pseudonym), einer 18-jährigen Frau mit afroamerikanischem Hintergrund.



Fallstudie: Foxy Boston

Zentral für die Diskussion wurden die **Charakteristika der Interviewerinnen:**

1. Interview. Geführt von Faye McNair-Knox, einer 41-jährigen afroamerikanischen Stanford-Dozentin; im Beisein ihrer 16-jährigen Tochter, die aus East Palo Alto stammt.

2. Interview. Geführt von Beth (Pseudonym), einer 25-jährigen euroamerikanischen Stanford-Doktorandin, die Foxy unbekannt war.

Unterschiede bzgl. des Alters, der Bekanntheit mit der Interviewten, des ethnischen Hintergrunds & der Anwesenheit eines *peers* (Alters- & Statusgenossin).



Fallstudie: Foxy Boston

Untersuchung typischer sprachlicher Merkmale des AAVE:

Nullkopula. *You crazy – You are crazy*

unflektiertes *be*. *He be working tomorrow – He is working tomorrow*

Nichtrealisierung des Pluralsuffixes *-s*. *they our friend – they are our friends*

... des Genitivsuffixes *-s*. *the teacher clerk – the teacher's clerk*

... des Flexionssuffixes *-s* (3. Pers. Sg.). *it seem like – it seems like*



Fallstudie: Foxy Boston

Untersuchte Merkmale gelten als charakteristisch für *casual speech* von AAVE-Sprecher:innen.

Beide Interviews unter sehr vergleichbaren Bedingungen: Keine anzunehmenden Unterschiede im Bezug auf die **Reflektiertheit** der Sprecherin.

Dennoch **signifikante *intraspeaker variation*** zwischen den Interviews: Höhere **Interaktionsaktivität** und **häufigere Realisierung** der AAVE-Merkmale im ersten Interview!



Fallstudie: Foxy Boston

| Variable | Foxy: Interview III (1990, afroamerikanische Interviewerin) | Foxy: Interview IV (1991, euroamerikanische Interviewerin) |
|--|---|--|
| Abwesenheit Genitiv- <i>s</i> | 67 % (6/9) | 50 % (5/10) |
| Abwesenheit Plural- <i>s</i> | 1 % (4/282) | 0 % (0/230) |
| Abwesenheit Flexion- <i>s</i> 3. Pers. Sing | 73 % (83/114) | 36 % (45/124)* |
| Abwesenheit Kopula <i>is/are</i> (Nullkopula) | 70 % (197/283) | 40 % (70/176)* |
| Unflektiertes <i>be</i> | 385 (= 241/h) | 97 (= 78/h)* |



Fallstudie: Foxy Boston

Die Reflektiertheit des Sprechens (*casual vs. careful speech*) kann diese Variation nicht erklären.

Viel eher findet hier **Akkommodation** statt, die Anpassung an den Sprachstil der anderen Person (bzw. die Vorstellungen davon).

Man kann sich dem Gegenüber sprachlich annähern, um Nähe aufzubauen (*convergence*); oder sich sprachlich entfernen, um Distanz auszudrücken (*divergence*).

Person der Interviewerin als ausschlaggebende Variable für die Variation von Foxy!



Beobachterparadoxon

Lest die Abschnitte 5.2.3-5 in Spitzmüller (2022, S. 135-137) zur Erhebung soziolinguistischer Forschungsdaten und arbeitet die Kernkonzepte heraus. Wie wird mit dem Beobachterparadoxon umgegangen?



Beobachterparadoxon

Es ist generell ratsam, bei Datenerhebungen die eigentliche Forschungsfrage zu **verschleiern**, da sie Einfluss auf das Verhalten der Gewährspersonen haben könnte.

Im sprachlichen Kontext ist das nicht ohne Weiteres machbar (schließlich können wir nicht einfach Leute ohne deren Einwilligung aufzeichnen).

Man versucht also, durch **vertraute Rahmenbedingungen** oder **emotionale Themen** eine möglichst unreflektierte und damit authentische Sprechweise hervorzulocken.

