

Carolin Dix

GAT2 trifft das dem International SignWriting Alphabet (ISWA): Ein neues System für die Transkription von Multimodalität¹

Transkripte sind ein wichtiges Handwerkszeug interaktionsanalytischer Forschung. Während es jedoch für die Verschriftung verbaler und vokaler Interaktionsressourcen bereits etablierte Konventionen gibt, fehlt ein solches einheitliches System für die Transkription visuell-leiblicher Ressourcen sowie multimodaler Aspekte. Der Beitrag stellt in diesem Zusammenhang ein neues Transkriptionssystem vor, das bestehende Konventionen, v.a. zur Transkription verbaler und vokaler Ressourcen, mit dem Zeicheninventar des International SignWriting Alphabet (ISWA) verknüpft. Dadurch entsteht eine standardisierte, forschungsübergreifende und dennoch an die Daten, den Transkriptionszweck und die jeweilige Forschungsfrage anpassbare Konvention für die Transkription von Videodaten, die keine verbale Beschreibung visueller Handlungen mehr notwendig macht.

GAT2 meets ISWA

Dieser Beitrag entwirft ein forschungsübergreifend nutzbares Transkriptionssystem, das das Zeicheninventar des ISWA (Sutton 2010, Parkhurst & Parkhurst 2008) mit den Konventionen für die Verbaltranskription des GAT2 (Selting et al. 2009) zusammenführt.

GAT2 kann mit den gängigen Textverarbeitungsprogrammen (z.B. Word) und ebenso spezieller Transkriptionssoftware (z.B. FOLKER, PRAAT, f4, Elan etc.) genutzt werden. Die Zeichen des ISWA sind z.B. über das Programm „SignMaker 2017“² abrufbar und dort als Bilddatei exportierbar. Das Programm erlaubt es zudem nicht nur, bereits bestehende Zeichen zu nutzen, sondern auch individuelle Zeichenkombinationen zu erstellen (siehe Beispiel 4). Dadurch kann das Repertoire des ISWA forschungsbezogen erweitert werden, um nichtstandardisierte visuell-leibliche Aspekte darzustellen. Eine Übertragung der ikonischen, indexikalischen und symbolischen Zeichen in Transkriptionsprogramme wie Elan ist bisher noch nicht möglich. Die Multimodaltranskripte in diesem Beitrag wurden daher in einem Textverarbeitungsprogramm erstellt.

Aufbau des Multimodaltranskripts

Aufgrund der temporalen Struktur nicht nur des gesprochenen Wortes, sondern auch des leiblichen Handelns, bildet die Repräsentation der zeitlichen Erstreckung einer Interaktion den Ausgangspunkt der Transkription. Dies kann sowohl verbale und vokale Ressourcen umfassen (das von einer Person verbal Geäußerte) als auch Stille oder Pausen, d.h. Interaktionsphasen, in denen nicht gesprochen wird.

¹ Dix, Carolin (forthcoming): GAT2 trifft das International SignWriting Alphabet (ISWA): Ein neues System für die Transkription von Multimodalität. In: Grawunder, Sven; Cordula Schwarze (Hrsg.): Transkription und Annotation gesprochener Sprache und multimodaler Interaktion. Konzepte, Probleme, Lösungen. Tübingen: Narr.

² Zugang zum Programm über: <http://www.signbank.org/signmaker/#?ui=de&dictionary=gsg> oder <https://slevinski.github.io/signmaker/#?ui=en>, zuletzt abgerufen am 30.10.2020; Stephen Slevinski stellt die technischen Details des Programms zudem in verschiedenen Videos vor (YouTube-Kanal: SignWriting Channel), siehe u.a. <https://www.youtube.com/watch?v=gQNHRBWDok8&t=5s>, zuletzt abgerufen am 11.04.2021

Nach der Konvention des Zeilentranskripts GAT2 gestaltet, wird jede Intonationsphrase auf einer eigenen, nummerierten Zeile notiert. Diese wird als Verbalzeile bezeichnet (Abb. 9).

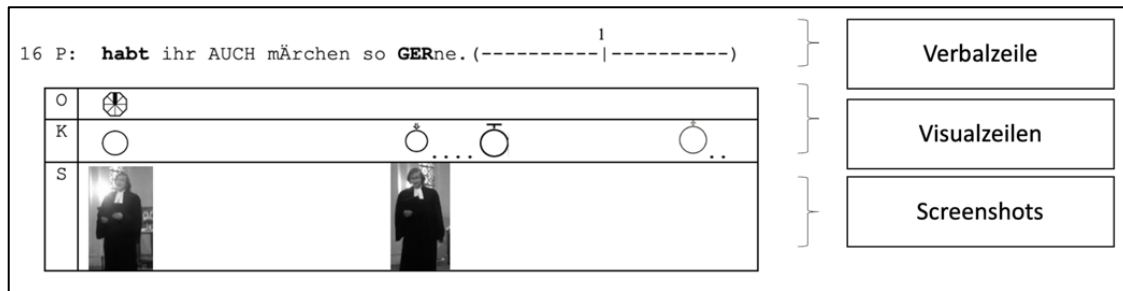


Abb. 1: Transkriptaufbau

Am Beginn einer Zeile wird die Sigle der Person vermerkt, der der Interaktionsbeitrag zugeordnet wird (in Abb. 9: P für Prediger). Phasen der Abwesenheit verbaler Äußerungen werden, abweichend von GAT2, nach Zehntelsekunden aufgeschlüsselt, wie Heath et al. es vorschlagen (Heath et al. 2010, Luff & Heath 2015: 371) und ebenfalls auf der Verbalzeile transkribiert. Ein senkrechter Strich mit einer darüberstehenden Zahl markiert eine volle Sekunde. Durch diese Art der Repräsentation kann genauer dargestellt werden, an welcher Stelle z.B. einer Pause eine Bewegung einsetzt, ihren Höhepunkt erreicht oder endet.

Unterhalb der Verbalzeile werden dann Visualzeilen eingefügt auf denen jeweils getrennt voneinander Phänomene unterschiedlicher Ressourcenebenen (z.B. Körperorientierung, Kopfposition, Handstellungen, Schulterbewegungen, faziale Gestik etc.) dargestellt werden. Dabei handelt es sich um eine künstliche Trennung, die im weiteren Datenanalyseprozess wieder aufgehoben werden muss. Die einzelnen Zeilen erhalten keine Nummerierung. So kann die Anzahl der Zeilen nicht nur an das jeweilige Datenmaterial angepasst, der Feinheitgrad und die Komplexitätsstufe des Transkripts variiert, sondern auch ein bereits vorhandenes Verbaltranskript nachträglich ergänzen werden. Unterhalb der Sprechersigle platziert, erhält jede Visualzeile eine Zeilensigle, die die transkribierte Ressourcenebene angibt. In den nachfolgenden Beispielen werden z.B. Zeilen für die Ausrichtung der Körper im Raum (Sigle O), die Bewegung des gesamten Körpers (Sigle M), die Bewegungen des Kopfes (Sigle K), der Augenlider (Sigle L), der Augenbrauen (Sigle E), des Oberkörpers/Torsos (Sigle T), Blickveränderungen (Sigle G) sowie Positionen und Bewegungen der Hände (Sigle H) notiert. Welche Zeilen notiert werden und ob darüber hinaus weitere angeführt werden, die hier nicht erwähnt sind, hängt vom jeweiligen Datenmaterial ab. Ein wesentlicher Bestandteil des Transkripts sind darüber hinaus unterhalb der Visualzeilen eingefügte Standbilder (Sigle S). Diese verdeutlichen das Zusammenspiel unterschiedlicher Ressourcen zu einem bestimmten Zeitpunkt der Interaktion und unterstützen die ganzheitliche Betrachtung der einzelnen Ebenen. Ein zusätzliches Visualisierungselement sind die entsprechenden Videoausschnitte, die, sofern dies aus rechtlichen Gründen möglich ist, z.B. über QR-Codes zur Verfügung gestellt werden. Dadurch entsteht eine enge Verzahnung zwischen dem Video und dem Transkript.

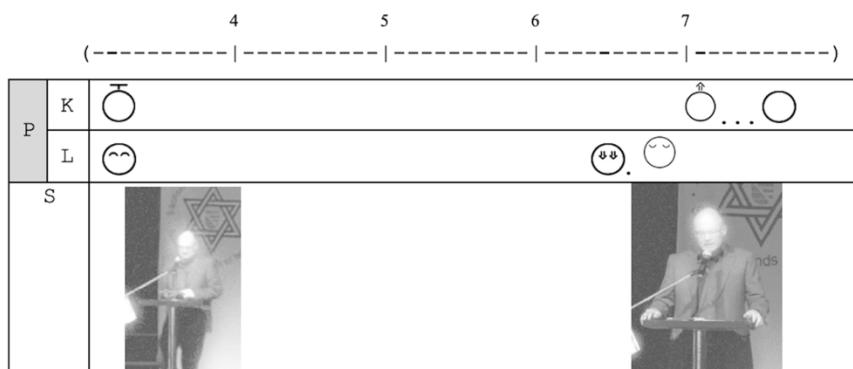
Innerhalb der einzelnen Visualzeilen werden die ISWA-Zeichen aus der Produzentenperspektive notiert (siehe Abschnitt 3) und exakt unterhalb der Stelle im Verbaltranskript vermerkt, an der sie parallel zum Gesprochenen bzw. in Relation zur vergehenden Zeit (bei Stille) auftreten. Um eine noch bessere Alignierung zu erreichen, kann die entsprechende Silbe oder Pause fett hervorgehoben werden, da diese Markierung in GAT2 noch nicht mit einer anderen Bedeutung belegt ist.

Transkription intrapersoneller Koordinierung, Kopfbewegungen und Mimik

Der nachfolgende Auszug zeigt den Beginn einer christlichen Predigt, der gänzlich ohne verbale Ressourcen realisiert wird. In der nummerierten Visualzeile wird aufgrund dessen die zeitliche Erstreckung in Zehntelsekunden aufgeschlüsselt. Der Fokus liegt auf dem Prediger (P) und den von ihm genutzten visuell-leiblichen Ressourcen, in diesem Beispiel exemplarisch die Position und Bewegung des Kopfes (Sigle K) und die Position der Augenlider (Sigle L). Diese Aspekte wurden auf jeweils einer eigenen Visualzeile notiert.

Beispiel 1: Gebet

01 P: (3.0) ((Gang zum Predigtort))



02 vater im himmel wenn (.) !DU! nicht;

Nach dem drei Sekunden dauernden Gang des Predigers (P) von seinem Sitzplatz zu dem auf einer Bühne stehenden Pult, von dem aus er die Predigt halten wird, tritt er mit Erreichen des Predigtortes in eine Phase der intrapersonellen Koordination ein. Der Prediger steht zunächst mit nach unten gesenktem Kopf (○) am Pult, die Augenlider sind geöffnet (☺). Durch die Nutzung des ISWA können an dieser Stelle des Beispiels nicht nur Bewegungen repräsentiert werden, sondern auch statische bzw. gehaltene Positionen aus denen heraus sich Bewegungen entfalten. Nach weiteren drei Sekunden in dieser Position (währenddessen er mit seinem Tablet beschäftigt ist), schließt der Prediger seine Augenlider (☹, Sek.6,5). Das ISWA Zeichen markiert den Beginn der Bewegung, die Dauer der Bewegung wird mithilfe von Punkten angezeigt. Dadurch lässt sich nicht nur die Temporalität einer einzelnen Ressource abbilden, sondern auch die Simultaneität mehrere Ressourcen in ihrer jeweiligen zeitlichen Erstreckung. Mit nun geschlossenen Augen (☹) hebt der Prediger seinen Kopf soweit an (○[↑]), dass er in eine gerade Position kommt (○). Um das Transkript trotz seiner Komplexität übersichtlich zu gestalten, werden die Zeichen für feste Positionen, hier z.B. die Kopfposition nach unten oder die geschlossenen Augen, jeweils am Anfang der Position notiert. Ein weiteres Zeichen wird erst dann eingefügt, wenn eine Bewegung einsetzt.

Besonders fruchtbar ist das hier vorgestellte System auch dort, wo bisherige Konventionen noch nicht ausgereift genug sind, wie bei der Transkription fazialer Gestik in der Interaktion (Alt 2020), etwa dem Heben der Augenbrauen. Im nachfolgenden Transkript handelt es sich um eine dyadische Sequenz innerhalb einer Multiparty-Interaktion im Rahmen eines Kochabends. Die Interaktion während des gemeinsamen Essens wurde mit einer 360°-Kamera aufgezeichnet. Person E (männlich) sitzt mit weiteren Personen am Tisch und schaut sich ein Kochbuch an, in dem Gerichte nach bekannten

Trickfilmen benannt sind. U.a. auch nach dem Film ‚Susi und Strolch‘. Zunächst verhandelt er mit Person TM (weiblich), wann der Film in die Kinos kam und richtet die Frage schließlich auch an Person A (weiblich), die an der Stirnseite des Tisches sitzt. Vier Ressourcenebenen werden im nachfolgenden Transkript dargestellt: die Position und Bewegung des Kopfes (Sigle K), die Position und Bewegung des Oberkörpers (Sigle T), die Position und Bewegung der Augenbrauen (Sigle E) und die Bewegung der Augenlider (Sigle L).

Beispiel 2: Susi und Strolch (Teil 1)

40 E: na der is ja auch schon URalt ne?
 41 TM: ja susi und STROLCH [ne?
 42 E: [susi und STROLCH? (---)

E	K	
	A	
A	K	
	T	
S		

43 A: WIE bitte?

A	K	
	T	
S		

44 E: in deiner JÜgend, (----)


A	E	
	K	
	L	
S		

Nachdem E zunächst mit Person TM über das Buch gesprochen hat, wendet er sich in Zeile 42 mit den Worten „susi und STROLCH?“ an A. Parallel mit seinem Redebeitrag, hebt er seinen Kopf aus einer nach unten gesenkten Position () an und dreht ihn nach links zu A (, Zeile 42), bis sein Kopf auf sie

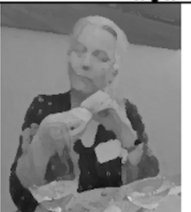
ausgerichtet ist (⊖). Auch A orientiert sich körperlich auf E, indem sie ihren Kopf in seine Richtung dreht (⊖, Zeile 42) und ihren Oberkörper nach rechts verlagert (→). Damit ratifiziert sie sich als Adressatin und stellt eine fokussierte Interaktion mit E her. Es folgt eine verbale Reparaturinitiation von A („WIE bitte?“, Z.43), auf die E jedoch nicht mit der Wiederholung seiner Äußerung reagiert, sondern mit einer *candidat answer*, wann A den Film zur Kenntnis genommen haben könnte („in deiner Jugend“, Z.44). Daraufhin re-etabliert A ihre Reparaturinitiation mithilfe visueller Ressourcen. Sie hebt auf die betont Silbe von „Jugend“ ihre Augenbrauen an (↑↑) und hält diese Position (⊖). Kurz darauf führt sie auch ihren Kopf nach oben (⊖) und schließt und öffnet in der folgenden Pause kurz beide Augenlider (⊖↕↕). Die Bewegung der Augenbrauen, der Augenlider und des Kopfes sind hier situierte visuell-leibliche Praktiken, um eine Reparatur zu initiieren.

Beispiel 3: Susi und Strolch (Teil 2)

45 susi und **STROLCH?**

A	E	⊖ .. ⊖
S		

46 wann WAR d_n das. (--)


A	T	← ...
K		↑↕↑
S		

47 SIEb[ziger?

48 A: [KENN ich. (---)

49 E: na vielleicht anfang ACHziger.

50 A: (-----)

A	G	⊖↕
E		↑↕ ⊖
S		

E wiederholt daraufhin zunächst den Gegenstand seiner Frage („susi und STROLCH?“, Z.45) und A senkt ihre Augenbrauen ab (⊖), sodass diese wieder eine neutrale Position erreichen (⊖). Während


der Frage „wann WAR d_n das“ (Z.46) von E nimmt A die Ausrichtung ihres Oberkörpers zurück, in dem sie ihre Schulterverlagerung aufhebt (←) und die Schultern in eine neutrale Position bringt (—). Parallel dazu bewegt sie ihren Kopf auf und ab (↑↓). Dieser als Nicken interpretierbaren Bewegung folgt die verbale Bestätigung „KENN ich“ (Z.48). Auf eine erneute *candidat answer* von E hinsichtlich des Erscheinungszeitraums des Films (Z.47 und 49) reagiert A schließlich mithilfe visueller Ressourcen, indem sie ihre Augenbrauen anhebt. Im Gegensatz zu Zeile 44 handelt es sich hier um ein Zucken, d.h. ein kurzes Heben und Senken (↕), Z.50, *eyebrow flash bzw. facial shrug*). A kontextualisiert so visuell-leiblich einen negativen epistemischen *stance* und damit Nichtwissen. Zudem wendet sie ihren Blick ab (⊙), Z.50) und hebt so die fokussierte Interaktion auf. Mithilfe der in das Transkript integrierten ISWA-Zeichen können die Art der fazialen Bewegung (kurzes Zucken oder länger gehaltene Position), ihre temporale Entfaltung und das multimodale Zusammenspiel mit verbalen und anderen visuellen Ressourcen, wie der Bewegung des Kopfes, ohne Bewegungsparaphrase repräsentiert werden. Das Heben der Augenbrauen wird dann als faziale Geste rekonstruierbar, die situativ und kontextsensitiv von den Interagierenden eingesetzt wird.

Transkription interpersoneller Koordinierung und körperlicher Ausrichtung


Beispiel 2 hat in Z.42 bereits gezeigt, wie das hier vorgestellte System dyadische Interaktionen verschriftet und damit nicht nur die intrapersonelle, sondern auch die interpersonelle Orchestrierung unterschiedlicher interaktiver Ressourcen deutlich gemacht werden kann. Das nachfolgende Beispiel aus einer Dialogpredigt zeigt hier die Möglichkeiten des Systems noch einmal auf.

Beispiel 4: Gedanken

226 P1: °h es beginnt IMMer;

P1	O	↻ ← ↻
	K	○
	H	■ ↑
P2	O	↻
	K	○
S		

227 mit_nem geDANken in deinem HERzen;

P1	O	↻ ↻
	M	↓
	T	— ↑ ?
	H	↻ ↻
P2	O	↻ ... ↻
	K	○ ... ○
S		

228 den du dort WOHNen lässt; (-)

Werden mehrere Personen in einem Transkript abgebildet, wird vor die einzelnen Visualzeilen die Sigle des jeweiligen Interagierenden platziert, um diese zuordnen zu können. Der weitere Aufbau des Transkripts ist identisch mit dem für Einzelpersonen. Das obige Beispiel zeigt einen Auszug aus einer Dialogpredigt, bei der sich zwei Prediger gleichzeitig am Predigtort befinden und gemeinsam den Wortbeitrag gestalten. Das Visualtranskript besteht aus drei Bereichen: Der Transkription des ersten Predigers (P1, links im Bild, Abb.10), der Transkription des zweiten Predigers (P2, rechts im Bild, Abb.10) und den Standbildern.




Abb. 2: Räumliche Anordnung Bühnenformat Predigt

Folgende Visualebenen sind im Transkript repräsentiert: Für P1 die Ebenen der Körperorientierung (Sigle O), des Kopfes (Sigle K), der Bewegung des gesamten Körpers (Sigle M), des Oberkörpers (Sigle T) und der Hände (Sigle H); für P2 die Ebenen der Körperorientierung (Sigle O) und der Kopfposition (Sigle K). Auf jeder dieser Ebenen wird nun aus Produzentenperspektive, aligniert mit dem Gesprochenen und/oder anderen visuellen Ressourcen, abgebildet, welche Positionen und Bewegungen eingenommen bzw. ausgeführt werden. Dadurch entsteht eine Darstellung der inter- und intrapersonellen Orchestrierung der genutzten Ressourcen. In diesem Beispiel spielt vor allem die körperliche Ausrichtung der beteiligten Personen eine wesentliche Rolle bei der Analyse (Sigle O). Da es sich hier um einen Auszug aus einer Predigt handelt, die prototypisch eine Interaktionsordnung als Bühnenformat aufweist, wird die Ausrichtung auf die Gemeinde als deiktische Verankerung, d.h. als ‚Vorne‘ der Prediger, festgelegt und folgendes Zeichen genutzt: \odot , um diese Position zu repräsentieren. Am Beginn des Beispielauszugs ist P1 schräg nach links ausgerichtet (\odot , Z.226), verändert diese Position dann, indem er sich zunächst nach links dreht (\curvearrowright), kurzzeitig eine Position face-to-face zu P2 einnimmt (\otimes), sich anschließend wieder nach rechts dreht (\curvearrowleft), bis er frontal auf die vor ihm sitzende Gemeinde ausgerichtet ist (\odot). Parallel zu dieser letzten Drehung geht er nach hinten (\downarrow). Da es sich hier um eine Bewegung des ganzen Körpers handelt, die nicht zwingend auch eine Veränderung in der körperlichen Orientierung beinhaltet, wird diese Positionsveränderung auf einer eigenen Visualzeile notiert. Der Kopf von P1 verharrt in dieser Zeit in einer geraden Position (\circ). Da sich daran im weiteren Verlauf seines Redebeitrags nichts ändert, wird diese Visualzeile nur am Anfang notiert und in der nächsten Intonationsphrase nicht wieder aufgenommen. Dies erhöht bzw. gewährleistet die Übersichtlichkeit des Transkripts. Während er spricht, verändert P1 nicht nur seine Körper-, sondern auch seine Handposition. Zunächst bewegt er auf die erste Silbe von „IMMER“ (Z.226) seine rechte Hand nach vorn (\uparrow). Der Zeigefinger ist dabei ausgestreckt, die Handfläche gekrümmt, der Handrücken zeigt nach oben (\blacksquare). Im Anschluss daran dreht er sein Handgelenk nach vorn (\curvearrowright), führt gleichzeitig die Hand zu sich und berührt schließlich mit der geöffneten Handfläche bei der betonten

Silbe von „geDANken“ (Z.227) seine Brust (☉). Da es sich dabei um eine komplexe Berührung handelt, wird das entsprechende Zeichen aus unterschiedlichen ISWA-Zeichen, angepasst an die Daten, individuell mithilfe des Programms SignMaker 2017 zusammengestellt. Das Zeichen repräsentiert dann nicht nur die Handposition (die Handfläche ist zum Prediger gerichtet, ☞), sondern auch die Stelle der Berührung (am Oberkörper, ☉) und den Berührungskontakt (markiert durch den Asterisk). Während P1 spricht, steht P2 zunächst seitlich nach rechts vorn ausgerichtet (☉, Z.226) und mit nach links unten gesenktem Kopf (☉) am Rednerpult. Erst bei „geDANken“ wendet er sich nach rechts (☞, Z.227) und verharret schließlich auf P1 ausgerichtet (☉). Während P2 in dieser Position bleibt, verändert P1 immer wieder seine körperliche Ausrichtung, sodass sowohl eine Formation face-to-face als auch face-to-side erkennbar wird. Gleichzeitig mit seiner körperlichen Bewegung hebt P2 den Kopf an (☉) und bringt ihn dadurch in eine gerade Position (☉), ebenfalls auf P1 ausgerichtet. Konzentriert sich die Forschungsfrage vor allem auf die Veränderungen in der körperlichen Ausrichtung der Prediger zueinander und zur Gemeinde, kann das Transkript z.B. zu Präsentationszwecken auf die entsprechenden Visualzeilen eingegrenzt werden (Beispiel 5).

Beispiel 5: Gedanken (nur körperliche Ausrichtung)

226 P1: °h es beginnt IMMer;

P1	O	☉ ☞ ☉
P2	O	☉
S		

227 mit_nem geDANken in deinem HERzen;

P1	O	☞ ☉
P2	O	☞ ... ☉
S		

228 den du dort WOHNen lässt; (-)
 229 und der dort (.) seine KREIse ziehen kann.

P1	O	☞ ☉
P2	O	☉
S		

Das Transkript kann so dazu beitragen die Positionsveränderungen der zwei Prediger während ihres Wortbeitrags nachvollziehbar zu machen und den Wechsel in der körperlichen Ausrichtung der Prediger zueinander und zur Gemeinde als wichtiges Element und den Wechsel zwischen Bewegungen

und statischen Positionen als charakteristischen Aspekt innerhalb von Dialogpredigten herauszuarbeiten.

Anhand der gezeigten Beispiele wird deutlich, dass das System unterschiedliche Ressourcenebenen berücksichtigt, die entsprechend des Materials und der Fragestellung angepasst werden können, sowohl statische Positionen als auch Bewegungen abbilden kann und die temporale Entfaltung einer Ressource genauso wie die Simultaneität unterschiedlicher Ressourcen, intra- wie interpersonell darstellt.

Carolin Dix

Universität Bayreuth, Universitätsstraße 30, 95440 Bayreuth

carolin.dix@uni-bayreuth.de

Forschungsschwerpunkte: Gesprächsforschung/ (multimodale) Interaktionsforschung, Sprache und Religion, Analyse kommunikativer Gattungen, Videoanalyse

<https://www.gl.uni-bayreuth.de/de/team/carolin-dix/index.php#tabPerson>