

Herausgegeben von Johannes Gerwien,  
Ines Marberg und Kristian Nicolaisen

# DIE KOGNITIVE PERSPEKTIVE

Wie Menschen über die Welt sprechen

Festschrift zu Ehren von Christiane von Stutterheim

HEIDELBERG  
UNIVERSITY PUBLISHING



# **Die kognitive Perspektive**

Wie Menschen über die Welt sprechen



Christiane von Stutterheim, kurz vor dem Wechsel 1984  
von Nijmegen nach Heidelberg

# **Die kognitive Perspektive**

Wie Menschen über die Welt sprechen


**Festschrift zu Ehren von  
Christiane von Stutterheim**


Herausgegeben von


**Johannes Gerwien, Ines Marberg  
und Kristian Nicolaisen**

HEIDELBERG  
UNIVERSITY PUBLISHING

**ORCID®**

Johannes Gerwien  <https://orcid.org/0000-0002-6207-8167>

Ines Marberg  <https://orcid.org/0000-0002-1101-7145>

Kristian Nicolaisen  <https://orcid.org/0000-0002-6195-0021>

**Bibliografische Information der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <https://dnb.dnb.de> abrufbar.



Dieses Werk ist unter der Creative-Commons-Lizenz CC BY-NC-SA 4.0 veröffentlicht. Die Umschlaggestaltung unterliegt der Creative-Commons-Lizenz CC BY-ND 4.0.

Publiziert bei Heidelberg University Publishing (heiUP), 2024

Universität Heidelberg / Universitätsbibliothek  
Heidelberg University Publishing (heiUP)  
Grabengasse 1, 69117 Heidelberg  
<https://heiup.uni-heidelberg.de>

Die Online-Version dieser Publikation ist auf den Verlagswebseiten von Heidelberg University Publishing <https://heiup.uni-heidelberg.de> dauerhaft frei verfügbar (Open Access).  
urn: urn:nbn:de:bsz:16-heiup-book-1292-1  
doi: <https://doi.org/10.17885/heiup.1292>

Text © 2024, das Copyright der Texte liegt bei den jeweiligen Verfasser:innen.

Umschlagabbildung: Susanne Palme-Waldemer, Stills (Detail), 2020, Fotografie und Malerei

ISBN 978-3-96822-163-2 (Softcover)

ISBN 978-3-96822-162-5 (PDF)

# Inhaltsverzeichnis

- Wolfgang Klein  
1 **Wie sollen wir leben?**
- Christine Dimroth, Cecilia Andorno, Sandra Benazzo  
23 **When Discourse Elicitation Tasks Go Dialogue:  
Introducing Entities in French, German, and Italian**
- Silvia Natale  
59 **The Use of Progressive Aspect in Two Varieties  
of Swiss German**
- Monique Lambert, Mary Carroll  
85 **Obstacles in Acquiring a Second Language:  
A Key to the Impact of Typological Differences  
on Conceptualization in Language Production**
- Hana Klages  
109 **„Es ist *der* Ritter, obwohl du ‚sie‘ gesagt hast“:  
Auflösung anaphorischer Personalpronomina im  
Erwerb des Deutschen als Erst- und Zweitsprache**
- Rolf Koeppel  
139 **Unbeabsichtigte Verursachung: Zu Widerständen  
beim Erwerb einer Konstruktion in der Fremdsprache  
Deutsch und Möglichkeiten ihrer Vermittlung**
- Zoriana Falinska  
165 **Man darf nicht verallgemeinern: Überlegungen  
zur Angemessenheit der (in)definiten NPn in  
generischen Kontexten**
- Michael Herweg  
195 **Varieties of Manner of Motion: A Frame-Semantic  
Account**
- Johannes Gerwien, Ines Marberg, Kristian Nicolaisen  
225 **What Are Events?**
- Monique Flecken  
257 **The Cognitive Perspective: Language and Cognition  
in Event Construal**
- 263 **Autorenverzeichnis**





---

Wolfgang Klein

# Wie sollen wir leben?

---

*In vollen Zügen!*

## 1 Incipit

Diese Frage hatte ich, ich weiß nicht mehr weshalb und wann genau, es war aber im Jahre 1984, auf einen kleinen Zettel geschrieben und an ein Buchregal geklebt, und die lakonische Antwort, von Christiane von Stutterheim daruntergeschrieben, stand kurz danach auf diesem Zettel. Es ist keine der drei Fragen, über die Immanuel Kant in der „Kritik der reinen Vernunft“ eingangs sagt (1781, Vorrede): „Die menschliche Vernunft hat das besondere Schicksal in einer Gattung ihrer Erkenntnisse: daß sie durch Fragen belästigt wird, die sie nicht abweisen kann, denn sie sind ihr durch die Natur der Vernunft selbst aufgegeben, die sie aber auch nicht beantworten kann, denn sie übersteigen alles Vermögen der menschlichen Vernunft.“ und die er dann gegen Ende der Kritik der reinen Vernunft nennt: „1. Was kann ich wissen? 2. Was soll ich tun? 3. Was darf ich hoffen?“ Aber wie diese belästigt sie uns das ganze Leben hindurch und die Vernunft kann sie nicht beantworten. Aber anders als für diese drei gibt es eine Antwort, nämlich die, die das Leben selbst erteilt. Und wie diese ist es eine Frage, die man am Ende anders stellt als am Anfang, denn da fragt man sich: „Was weiß ich nun? Was hätte ich tun sollen? Was von dem Erhofften ist eingetreten? Habe ich gelebt, wie ich hätte leben sollen?“

Derlei geht mir durch den Sinn, wenn ich über die fast vier Jahrzehnte nachdenke, die seit diesem in vielerlei Hinsicht prägenden Jahr vergangen sind und in denen Christiane von Stutterheim und ich in Wissenschaft und Freundschaft verbunden waren. So will ich im Folgenden einige Gedanken an jene Zeit mit einigen verbinden, die sich daran anknüpfen.

## 2 1984 – 1

Im April jenes Jahres wurde sie mit ihrer von Norbert Dittmar und mir betreuten Dissertation *Temporalität in der Zweitsprache. Eine Untersuchung zum Erwerb des Deutschen durch türkische Gastarbeiter* promoviert; die Arbeit, von der FU Berlin preisgekrönt, ist zwei Jahre später als Buch erschienen.

Entstanden ist sie im Umkreis des europaweiten und vom Max-Planck-Institut für Psycholinguistik geleiteten „European Science Foundation Projects“, und wie dieses befasst sie sich mit dem natürlichen, d. h. nicht durch Unterricht gesteuerten Erwerb einer Zweitsprache durch erwachsene Lerner im alltäglichen Kontakt. So teilt sie mit diesem Projekt auch einige Grundannahmen, beispielsweise dass die Basis authentische Sprachdaten bilden, die verschiedene Grade der Sprachbeherrschung in der tatsächlichen Kommunikation widerspiegeln. Was die Arbeit im Besonderen auszeichnet, wird aus der einleitenden Formulierung der Ziele deutlich. Da heißt es (von Stutterheim, 1986, S. 1):

Jeder Mensch bildet im Rahmen seiner kognitiven Entwicklung bestimmte zeitliche Konzepte aus. Die besondere Ausprägung der Zeitkonzepte, über die ein erwachsener Sprecher verfügt, wird im wesentlichen im Verlauf des Erstspracherwerbs festgelegt. Die Sprachen unterscheiden sich in bezug auf die in ihnen gefaßten zeitlichen Kategorien. Für einen Sprecher werden durch die Muttersprache bestimmte begriffliche Kategorien in den Vordergrund von Konzeptualisierungsvorgängen gerückt.

So wird beispielsweise der Sprecher einer Sprache, die drei Stufen der Vergangenheit obligatorisch markiert (z. B. ein Tag vor dem Sprechzeitpunkt, eine Woche davor, mehr als eine Woche davor), diese Unterscheidung für jedes berichtete Ereignis der Vergangenheit treffen, während der Sprecher einer anderen Sprache, die nur eine allgemeine Vergangenheit grammatisch obligatorisch ausdrückt, diese Unterscheidung zwar prinzipiell vornehmen kann, aber nicht für alle vergangenen Ereignisse vornehmen wird. Das Verhältnis von sprachlicher und konzeptueller Repräsentation wird im folgenden Gegenstand genauer Betrachtung sein.

Zunächst soll allgemein festgehalten werden, dass der erwachsene Lerner mit einem ausgebildeten, durch seine Muttersprache geprägten Konzept von Temporalität in den Zweitspracherwerbsprozess eintritt. Verläuft der Prozess ohne den Einfluss eines von außen steuernden Unterrichts, so muss der Lerner selbst Kriterien besitzen, durch die die selektive Aufnahme von Formen der Zielsprache bestimmt wird. Die vorliegende Arbeit beruht auf der Hypothese, daß die konzeptuellen Strukturen eine zentrale Rolle bei dieser Form des Zweitspracherwerbs und -gebrauchs spielen. Am Beispiel des ungesteuerten Spracherwerbs türkischer Gastarbeiter soll untersucht werden, in welchen Formen temporale Referenz sprachlich repräsentiert wird und welche Konzepte diesen Formen zu Grunde liegen.

Ich zitiere das hier aus zwei Gründen so ausführlich. Zum einen formuliert es klar einen Gedanken, der für jedes Verständnis des Spracherwerbs – und damit auch des Zweitspracherwerbs Erwachsener – fundamental ist: Eine Sprache zu lernen bedeutet nicht, ihre lexikalischen und grammatischen Formen und Regeln zu lernen; es bedeutet, Konzepte – also Gedanken, Gefühle, Wünsche, was uns halt so alles durch den im Kopf geht – in wahrnehmbare Zeichen – Lautfolgen oder Graphemfolgen – umzusetzen und umgekehrt aus diesen zu erschließen. Der Sprachunterricht ist nur ein Versuch, in einen naturgegebenen Lernprozess einzugreifen und ihn zu optimieren. Dazu muss man aber verstehen, welchen Prinzipien dieser Prozess folgt; das wiederum ist nur möglich, wenn man untersucht, wie er *in vivo* funktioniert. So vorzugehen ist freilich sehr viel aufwendiger und sehr viel schwieriger als auszuzählen, welche und wie viele Fehler die Lerner beim gesteuerten Spracherwerb machen; so nimmt es denn nicht wunder, dass auch heute bei weitem die meisten Veröffentlichungen zum Zweitspracherwerb seiner gesteuerten Form gewidmet sind. Und es nimmt nicht wunder, dass es kaum einen Nachweis dafür gibt, dass trotz aller darum bemühter Forschung heute eine zweite Sprache im Unterricht besser gelernt würde als vor einem halben Jahrhundert.

Zum andern zeigt das Zitat, wie weit diese Dissertation im Kern bereits die Leitfrage eines wesentlichen Teils ihrer späteren Forschung benennt: Wie hängen Konzeptualisierung und Formulierung in den einzelnen Sprachen zusammen?

### 3 1984 – 2

Im Herbst jenes Jahres hat Christiane ihre Tätigkeit an der Universität Heidelberg aufgenommen. Das war in gewisser Weise ein Bruch, in gewisser Weise eine Fortsetzung, wie es immer im Leben ist, auch wenn man es selber nicht immer merkt. Es war ein Bruch insofern, als die Beschäftigung mit dem Zweitspracherwerb Erwachsener in den Hintergrund getreten, wenn auch nicht verschwunden ist (siehe etwa Klein und von Stutterheim, 1986). Eine Fortsetzung war es, weil weiterhin „[d]as Verhältnis von sprachlicher und konzeptueller Repräsentation ... Gegenstand genauerer Betrachtung“ ist – allerdings nunmehr vor allem im Vergleich verschiedener Sprachen. Das ist in neuer Form die alte Frage nach dem Zusammenhang zwischen Sprache und Denken, über die die Philosophen schon zweieinhalbtausend Jahre lang nachdenken. Auf diese gedankliche Arbeit kann man allerdings das Diktum in Anwendung bringen, mit dem Immanuel Kant die zweite Ausgabe der „Kritik der reinen Vernunft“ (1787, Vorrede) eröffnet:

Ob die Bearbeitung der Erkenntnisse, die zum Vernunftgeschäfte gehören, den sicheren Gang einer Wissenschaft gehe oder nicht, das läßt sich bald aus dem Erfolg beurteilen. Wenn sie nach viel gemachten Anstalten und Zurüstungen, so bald es zum Zweck kommt, in Stecken gerät, oder, um diesen zu erreichen, öfters wieder zurückgehen und einen andern Weg einschlagen muß; imgleichen wenn es nicht möglich ist, die verschiedenen Mitarbeiter in der Art, wie die gemeinschaftliche Absicht erfolgt werden soll, einhellig zu machen: so kann man immer überzeugt sein, daß ein solches Studium bei weitem noch nicht den sicheren Gang einer Wissenschaft eingeschlagen, sondern ein bloßes Herumtappen sei, ...

Nun ist die Bearbeitung von Erkenntnissen durch Philosophen eine andere als die der Sprachwissenschaftler, wenn es um Sprachliches geht. Man sieht es plastisch an der Art, wie Georg Wilhelm Hegel das Jetzt analysiert (Phänomenologie des Geistes, Abschnitt A,1):

Auf die Frage: *was ist das Jetzt* antworten wir also zum Beispiel: *das Jetzt ist die Nacht*. Um die Wahrheit dieser sinnlichen Gewißheit zu prüfen, ist ein einfacher Versuch hinreichend. Wir schreiben diese Wahrheit auf; eine Wahrheit kann durch Aufschreiben nicht verlieren; ebensowenig dadurch, daß wir sie aufbewahren. Sehen wir *jetzt, diesen Mittag*, die aufgeschriebene Wahrheit wieder an, so werden wir sagen müssen, daß sie schal geworden ist.

Das Jetzt, welches Nacht ist, wird *aufbewahrt*, d. h. es wird behandelt als das, für was es ausgegeben wird, als ein *Seiendes*; es erweist sich aber vielmehr als ein Nichtseiendes. Das *Jetzt* selbst erhält sich wohl, aber als ein solches, das nicht Nacht ist; ebenso erhält es sich gegen den Tag, der es *jetzt* ist, als ein solches, das auch nicht Tag ist, oder als ein *Negatives* überhaupt. Dieses sich erhaltende Jetzt ist daher nicht ein unmittelbares, sondern ein vermitteltes; denn es ist als ein bleibendes und sich erhaltendes *dadurch* bestimmt, daß anderes, nämlich der Tag und die Nacht, nicht ist. Dabei ist es eben noch so einfach als zuvor.

Ein Sprachwissenschaftler würde hier sagen, dass das Wort *jetzt* ein Zeitintervall bezeichnet, das die jeweilige Sprechzeit enthält.

Nun haben sich Sprachwissenschaftler bis zu Hegels Tagen und noch lange danach mit wenigen Ausnahmen kaum für das Thema Sprache und Denken interessiert, und die beiden berühmtesten unter den Ausnahmen, nämlich Wilhelm von Humboldt und Benjamin Lee Whorf, waren eigentlich Hobbylinguisten. Insbesondere gab es vor 40 Jahren kaum empirische Untersuchungen, die über einige wenige Eigenheiten des Wortschatzes oder der Flexionsmorphologie hinausgegangen wären. Überhaupt nicht untersucht wurde, wie

Sprecher verschiedener Sprachen vorgehen, wenn sie ein und dieselbe kommunikative Aufgabe lösen sollen: Wie wird die Aufgabe konzeptualisiert, dies vor allem in Bezug auf fundamentale kognitive Kategorien von Raum und Zeit, und wie wird sie in die konkreten sprachlichen Formen der jeweiligen Sprache umgesetzt? Inwieweit hängt ersteres von dem ab, was letztere dafür zur Verfügung stellt – nicht nur von dem Potenzial, sondern auch von den Zwängen, die damit einhergehen? Dies sind in etwas ausbuchstabierter Form genau die Fragen, die im mittleren der drei zitierten Abschnitte aus der Dissertation am Beispiel der Vergangenheit illustriert wird. In Wirklichkeit ist die kommunikative Aufgabe sehr viel komplexer, sobald man über einfache Äußerungen hinausgeht und die Bildung von Texten betrachtet. Das „Heidelberger Projekt“ um Christiane von Stutterheim war nicht das einzige, das in den vergangenen Jahrzehnten den Zusammenhang zwischen sprachnahe Denken und dem tatsächlichen Sprachgebrauch vergleichend analysiert hat; aber die dortigen Untersuchungen sind die mit weitem Abstand systematischsten und differenziertesten, nicht zuletzt in ihrer Verbindung von eher traditionell-linguistischen und neueren psycholinguistischen Methoden. So ist ein Weg gezeigt.

#### 4 1984 – 3

Im Spätherbst jenes Jahres 1984, in derselben Wohnung, in der noch der Zettel am Buchregal hing, haben wir uns einmal über die Frage unterhalten, wieso sich die beiden fundamentalen Kategorien der menschlichen Kognition, nämlich Raum und Zeit, in Erzählungen und darüber hinaus in Texten überhaupt ganz unterschiedlich verhalten. Erzählungen folgen einem ganz fundamentalen Prinzip, das bereits die alten Griechen unter dem Begriff des (unzulässigen) „hysteron proteron“ gekannt haben: Das Spätere soll nicht als Erstes kommen. In neuerer Zeit wurde es in verschiedenen Weisen formuliert, von William Labov beispielsweise als das „Principle of Chronological Order“ oder Eve Clark als Regel „Order of mention follows order of events“. Der Grundgedanke ist immer derselbe: Die Zeit, über die geredet wird, verschiebt sich im Text von Satz zu Satz als ein „und dann“. Es ist ein Default-Prinzip; Abweichungen von der Regel sind durchaus möglich und werden gegebenenfalls markiert; wenn nicht, entstehen bestimmte rhetorische Effekte. Wieso gilt dies nicht gleichermaßen für den Raum? Da ist es nämlich umgekehrt: Der Raum, über den geredet wird, bleibt gleich – es sei denn, es wird eine Verschiebung markiert. Ebenso muss bei Personen, insbesondere einem Agens, eine Änderung im Regelfall markiert werden. Allerdings ist die Situation hier etwas komplizierter; manchmal muss aus grammatischen Gründen Beibehalten statt Verschieben markiert werden, etwa durch ein an deskriptivem Gehalt armes anaphorisches Element. Beobachtungen dieser Art legen den

Gedanken einer „referentiellen Verschiebung“ innerhalb der Sätze eines Textes nahe, die bestimmten Regeln folgt; wie diese Regeln aussehen, hängt einestails von den grammatischen Regularitäten der betreffenden Sprache ab, andernteils von der Art des Textes; diese wiederum hängt von der impliziten oder expliziten Frage ab, die der Text in seiner Gesamtheit beantwortet. Unsere ersten Überlegungen dazu haben wir 1985 in einem Aufsatz unter dem Titel *Text structure and referential movement* zu Papier gebracht, jedoch noch nicht veröffentlicht. In diesen seligen Zeiten, in denen Impact Factor und h-index unbekannte Größen waren, hat man seine Aufsätze nicht gleich auf den Markt gebracht, sondern hektographiert oder in Fotokopie an befreundete Kollegen geschickt und um Kommentare, Kritik und Beifall gebeten. Dies ist dann in einen stark überarbeiteten Aufsatz eingegangen, der 1987 unter dem Titel *Quaestio und referentielle Bewegung in Erzählungen* erschienen ist. Andere folgten, teils mit, teils ohne empirische Analysen. Die mit Abstand wichtigste Arbeit ist jedoch die 1997 erschienene Habilitationsschrift *Einige Prinzipien des Textaufbaus. Empirische Untersuchungen zur Produktion mündlicher Texte*, in der der Ansatz systematisch ausgearbeitet und auf mehrere Texttypen angewandt wurde. Es fällt mir schwer zu glauben, dass das auch schon über ein Vierteljahrhundert her ist.

Das Besondere an diesem Ansatz ist, dass er eine ganz alte Idee, die der Quaestio, mit einer neuen, der referentiellen Bewegung, verbunden hat. Im Folgenden will ich zunächst auf diese Verbindung und ihren historischen Hintergrund eingehen; daran schließen sich einige neue Überlegungen an, die sich aus der Idee der Quaestio und, damit verbunden, der Idee der von ihr festgelegten Alternativen ergeben.

## 5 Quaestio

Der Begriff Quaestio stammt aus der antiken Rhetorik, und wie bei so vielen Begriffen, die sich die Alten vor mehr als zwei Jahrtausenden ausgedacht haben, ist nicht ganz einfach zu verstehen, was genau sie damit gemeint haben. Ungenau bezieht er sich primär auf die Rede vor Gericht und bezeichnet den Redesachverhalt, der geklärt werden muss. Dabei kann es um zweierlei gehen, nämlich um den Tatbestand, den man durch Zeugen oder aus Indizien ermitteln muss, oder um die juristische Würdigung, die man aus dem Gesetz herleiten muss; dementsprechend unterscheidet man auch zwischen der *quaestio facti* und der *quaestio juris*. Letztlich hat sich dieser Gedanke bis heute in der Jurisprudenz gehalten, beispielsweise in der Unterscheidung zwischen Berufung und Revision; bei letzterer wird der Tatbestand als gegeben angenommen, geprüft wird nur noch auf mögliche Rechtsfehler.

Allerdings ist der alte Begriff der Quaestio nicht auf juristische Probleme beschränkt. Quintilian, dem wir die letzte antike Synthese der Redekunst verdanken, unterscheidet im Anschluss an frühere Rhetoriklehrer Rechtsfragen von Sachfragen; seltsamerweise ist die Unterscheidung daran festgemacht, ob die Frage geschrieben vorliegt oder nicht („In scripto sunt de iure, in non scripto de re“, Kapitel III, Abschnitt 5,4 ff). Bei beiden wird unterschieden zwischen definiten und indefiniten Fragen (quaestiones finitae und quaestiones infinitae); erstere beziehen sich auf bestimmte Personen, Zeiten oder Orte, letztere nicht: *Soll man heiraten?* ist eine quaestio infinita, die man durch generelle Überlegungen beantworten kann, beispielsweise wie Kierkegaard: „Heirate, und du wirst es bereuen. Heirate nicht, und du wirst es bereuen.“ Hingegen ist *Soll ich Cato heiraten?* eine quaestio finita, bei der andere Beweggründe in die Antwort eingehen, beispielsweise der von Pascal beschriebene Umstand „Le cœur a ses raisons que la raison ne connaît point“. Die Unterscheidung zwischen quaestio finita und quaestio infinita ist der einzige mir bekannte Fall aus der antiken Rhetorik, in dem bei einer begrifflichen Unterscheidung die konkrete sprachliche Form beigezogen wird. Das sollte naturgemäß auch Konsequenzen für die konkrete sprachliche Form der möglichen Antworten haben; darüber allerdings wird meines Wissens an keiner Stelle der antiken Rhetorik etwas gesagt, wie denn überhaupt der sprachliche Zusammenhang zwischen Frage und Antwort nirgends thematisiert wird.

Das ist in gewisser Weise anders in der Scholastik, die sich ja nicht nur auf die Heiligen Schriften stützt, sondern auch auf die Heiden unter den griechischen und römischen Autoren, wo es passt. Dort sind die Quaestio und die Art und Weise, wie sie beantwortet wird, das fundamentale Organisationsprinzip der Lehre.<sup>1</sup> Die Quaestio legt fest, welcher Redegegenstand zur Entscheidung ansteht; das sind keine Rechtsfragen, sondern Fragen des christlichen Glaubens. Bei der Antwort werden zunächst jene Argumente vorgebracht, die gegen die richtige Antwort sprechen; dann folgt, was gegen diese Einwände spricht, und abschließend wird die richtige Antwort im Detail ausgeführt. Hier ist ein Beispiel aus dem bedeutendsten theologischen Werk der Scholastik, der heute noch für Katholiken maßgeblichen *Summa theologiae* des heiligen Thomas von Aquin. Es geht in dieser Quaestio LII, pars prima, nicht um die berühmte spitzfindige Frage, wie viele Engel auf einer Nadelspitze Platz haben – diese Frage war in der Scholastik

---

1 Solche Organisationsprinzipien sind allerdings keine Erfindung der Scholastiker; Thomas von Aquin führt sie auf Boethius zurück. In Bezug auf Rechtsfragen findet sich Ähnliches beispielsweise bereits in Papinians *Quaestiones* (um 200 n. Chr.). Siehe dazu Babusiaux (2011) und allgemein zur Quaestio als scholastische Methode Hoye (1997).

kein relevantes Thema –, sondern um die Frage, auf wie vielen Nadeln ein Engel gleichzeitig Platz haben kann, allgemeiner um das Verhältnis von Engeln und Orten.<sup>2</sup>

Die Quaestio wird einleitend in drei Teilfragen aufgespalten, die dann separat im vorgegeben Schema beantwortet werden:

Deinde quaeritur de loco Angeli. Et circa hoc quaeruntur tria. Primo, utrum Angelus sit in loco. Secundo, utrum possit esse in pluribus locis simul. Tertio, utrum plures Angeli possint esse in eodem loco. [Nunmehr wird nach dem Ort eines Engels gefragt. Und dabei wird nach dreierlei gefragt. Erstens, ob ein Engel an einem Ort ist. Zweitens, ob er an mehreren Orten gleichzeitig sein kann. Drittens, ob mehrere Engel an demselben Ort sein können.]

Die Antwort auf die erste Frage, ob ein Engel an einem Ort ist, lautet wie folgt (die Übersetzung stammt von Celsus Maria Schneider; dieses und das obige Zitate nach der zweisprachigen Ausgabe unter <https://bkv.unifr.ch/de/works/sth/versions/summe-der-theologie>):

Erster Artikel. Jeder Engel ist in einem Orte.

a) Dagegen spricht: I. Boëtius sagt: „Allgemein ist es von den Weltweisen angenommen, daß unkörperliche Wesen in einem Orte nicht sind“ (lib. de hebdom.). Und Aristoteles (4 Phys.): „Nicht alles, was ist, findet sich an einem Orte, sondern nur der bewegliche Körper.“

II. Ein Ort ist nichts Anderes wie ein Umfang, der eine gewisse Lage hat. Alles also, was in einem Orte ist, hat eine gewisse Lage. Die Substanz des Engels aber ist frei von allem Umfange; also ist sie in keinem Orte.

III. Sein in einem Orte ist dasselbe, wie vom Orte umschlossen und von ihm gemessen sein. Das kann aber dem Engel nicht zukommen, daß sein Maß etwas Körperliches ist.

---

2 Ich habe diese Quaestio ausgesucht, weil sowohl Christiane von Stutterheim wie ich wiederholt zu Problemen der Raumreferenz veröffentlicht haben, dabei waren unsere Auffassungen nicht immer dieselben. Ich denke noch immer dankbar daran, wie sie seinerzeit auf einige Probleme aufmerksam gemacht hat, die sich einstellen, wenn es nicht um die Lokalisierung von Objekten, sondern die von Sachverhalten geht, also nicht *Die Tasse stand auf dem Tisch*, sondern beispielsweise *Unter dem Dach war es kalt*. Da wäre es nun interessant zu wissen, wie sich dies bei Entitäten verhält, die weder materielle Objekte noch Sachverhalte sind. Ich selber finde die Argumentation des heiligen Thomas schwer widerleglich.



Auf der anderen Seite heißt es im Gebete: „Deine heiligen Engel mögen wohnen in diesem Hause und uns im Frieden beschützen.“

b) Ich antworte, daß der Engel in einem Orte sich findet. Nicht aus demselben Grunde aber wird vom Engel dies ausgesagt wie vom Körper. Denn der Körper ist dadurch in einem Orte, daß er kraft der Berührung seines Umfanges an den Ort gebunden wird. Der Engel aber hat in seiner Natur keinen Umfang; wohl aber besitzt er einen gewissen Umfang in seiner wirkenden Kraft. Daher demgemäß weil der Engel seine Kraft auf einen Ort richtet, wird von ihm gesagt, er sei in einem körperlichen Orte. Danach wird also der Engel nicht gemessen vom Orte und er hat keine bestimmte Lage in einem Orte; das kommt nur einem Körper zu, der mit seinem Umfange den Ort berührt. Vielmehr mißt der Engel und hält zusammen durch seine Kraft den Körper, der im Orte ist und nicht umgekehrt. So ist ja auch die menschliche Seele im Körper; nicht als von diesem gemessen und zusammengehalten, sondern als ihn messend und zusammenhaltend.

In der Scholastik wird also klar vorgegeben, wie die Antwort auf eine strittige Frage auszusehen hat, auch wenn diese Vorgaben nicht immer so rigide befolgt werden. In ganz ähnlicher Weise wird auch die scholastische Disputation gehandhabt, die ja ein wesentliches Element der akademischen Ausbildung ausgemacht hat (manchmal wünscht man es sich zurück). Was hingegen nicht vorgegeben wird, ist die konkrete sprachliche Form der Antwort, ganz egal, ob sie in einem kurzen Satz erfolgte (was in der Regel nicht der Fall war) oder in einem mehr oder minder langen Text. Ein solcher Text war in sich sehr wohl strukturiert, aber nicht nach sprachlichen Kriterien.

Jetzt mache ich einen langen Sprung nach Nijmegen im Jahr 1978 (immer noch ein Drittel kürzer als der Sprung von Quintilian zu Thomas von Aquin). In diesem Jahr hatte der Soziologe Max Miller, damals Mitarbeiter an der Max-Planck-Projektgruppe Psycholinguistik, die schöne Idee, die Entwicklung moralischer Normen von Kindern nicht zu untersuchen, indem man sie einzeln befragt (da sagen Kinder gerne, was ihrer Einschätzung nach die Erwachsenen hören wollen), sondern dadurch, dass man sie unter sich in einer kleinen Gruppe über moralische Dilemmata diskutieren lässt, beispielsweise *Darf ein Mann eine Bank überfallen, wenn er nur so das Geld für die lebensrettende Operation seiner Frau beschaffen kann?*. Da sieht man, wie die Kinder tatsächlich ticken. Es zeigt sich, dass Kinder schon vor der Schulzeit erstaunlich gut mit einer solchen Quaestio infinita umgehen können, auch wenn man sich bisweilen über ihre Argumente wundert („Ja, sonst muss er sich doch eine neue Frau suchen“). Kollektive Argumentationen sind in der Argumentationstheorie so gut wie nie behandelt worden, erst recht

nicht solche unter Kindern; deshalb ist eine Analyse solcher authentischer Gesprächsdaten sehr schwierig. Max Miller hat mich damals gefragt, ob ich eine Idee dazu hätte, und so habe ich mir ein Verfahren ausgedacht, wie man zwei zusammengehörige, aber klar verschiedene Dinge analysieren kann: einerseits die „Logik der Argumentation“, d. h. die Art und Weise, wie das Argument im Hin und Wider entwickelt wird, und andererseits die „Logik des Arguments“, d. h. einer abstrakten propositionalen Struktur, in der eine Proposition durch andere gestützt wird, bis durch die letzte und oberste Proposition die Quaestio beantwortet wird. Das Ergebnis dieser Zusammenarbeit war ein kleiner gemeinsamer Aufsatz (Miller und Klein, 1981), ein ganzes Buch dazu von ihm und mehrere Aufsätze von mir, in denen das Verfahren entwickelt und auf Argumentationen unter Erwachsenen angewandt wird. Ein wesentliches Element darin war die Idee der „komplexen sprachlichen Handlung“, die kollektiv oder individuell sein kann. Argumentationen sind nur ein Sonderfall davon, es gibt viele andere, und eine jede kann durch eine charakteristische Frage – eine Quaestio – gekennzeichnet werden: *Soll man heiraten? Was spricht dagegen, bei minderen Vergehen die Prügelstrafe wieder einzuführen? Wo geht es denn hier zum Goethehaus? Wie spielt man Skat? Was ist denn mit deiner Nase passiert? Wo kommen die kleinen Kinder her?* und viele andere mehr. So erhält man eine sehr flexible Charakterisierung von Texttypen, die von mehreren gemeinsam oder von jemandem allein geschaffen werden. Damals habe ich einige solcher komplexen sprachlichen Handlungen untersucht, neben Argumentationen auch Wegauskünfte und Erzählungen; an der durch die Quaestio *Wie spielt man Skat?* initiierten Aufgabe bin ich elendiglich gescheitert.

Dieser Quaestio-Begriff geht insofern deutlich über den antiken und mittelalterlichen hinaus, als er erstens in flexibler Weise auf viele komplexe sprachliche Handlungen anwendbar ist, und zweitens dadurch, dass sich das Vorgehen und damit der zu schaffende Text weithin aus der Art der Quaestio ergibt. Die klassische Quaestio macht auch Vorgaben, sie sind aber sehr global, und sie strukturieren die sprachliche Gestalt der Antwort nicht. In meinen eigenen Arbeiten zu komplexen sprachlichen Handlungen war dies allerdings auch noch nicht berücksichtigt. Geleistet wurde es erst durch die Verbindung *Quaestio und referentielle Bewegung* – besser gesagt, partiell geleistet, denn die Vorgaben schränken nur den Spielraum ein, und es ist möglich, gegen sie zu verstoßen. Dies führt dann zu speziellen Effekten. In Klein und von Stutterheim (1987, S. 163) heißt es dazu einleitend:

Der Gesamtaufbau eines Textes, die Art und Weise, wie sich das Mitgeteilte von einer Äußerung zur nächsten entfaltet, schließlich auch der Aufbau der einzelnen Äußerung unterliegen einer Reihe von Beschränkungen. Viele darunter rühren daher, daß der Text in seiner Gesamtheit

dazu dient, eine – explizite oder implizite – Frage zu beantworten – die Quaestio des Textes. Die einzelnen Äußerungen liefern dazu in wohlhabender Weise jeweils einen bestimmten Beitrag.

Dieser Grundgedanke wird dann in fünf des längeren ausgeführten Punkten entfaltet; hier sind diese Punkte nur genannt, auf einen davon gehe ich im nächsten Abschnitt näher ein:

- A. Ein wohlgeordneter Text beantwortet eine Quaestio.
- B. Die Quaestio legt Hauptstruktur (Vordergrund) und Nebenstrukturen (Hintergrund) des Textes fest.
- C. Die Quaestio legt teilweise fest, was zur Topik und was zum Fokus einer Äußerung der Hauptstruktur zählt.
- D. Die Quaestio legt teilweise fest, wie bestimmte Referenzbereiche innerhalb einer Äußerung zu belegen sind.
- E. Die Quaestio legt teilweise die referentielle Bewegung zwischen Äußerungen der Hauptstruktur fest.

A konstatiert den Grundgedanken, den man als eine Entfaltung der antiken und mittelalterlichen Quaestio-Idee sehen kann; die folgenden Punkte fassen diese Entfaltung zusammen. B dient dazu, Äußerungen, die nicht direkt auf die Quaestio reagieren, wohl aber kommunikativ wichtig sein können, gleichsam auszusortieren; C greift nun in die einzelne Äußerung ein und legt dort (partiell) die Informationsstruktur fest; D legt partiell die semantischen Bestandteile, die insgesamt eine Äußerung ausmachen, fest, und E schließlich legt fest, wie diese semantischen Bestandteile sich von einer Äußerung zur nächsten weiterentwickeln; es sei an das Ausgangsbeispiel vom Herbst 1974 erinnert, dass sich die Zeit im Regelfall verschiebt, wenn nichts anderes gesagt wird, dass aber der Raum erhalten bleibt, wenn nichts anderes gesagt wird.

All dies wurde seinerzeit ausführlich konkretisiert. Manches würde man heute sicher anders machen, anderes wurde nie ausgenutzt. Im Folgenden will ich einen Punkt herausgreifen und weiterentwickeln. Er betrifft den notorisch schwierigen Begriff „Topik“, einen der Kernbegriffe der Informationsstruktur.

## 6 Alternative und Alternanten

Der heilige Thomas hat, wie wir gesehen haben, seine Quaestio LII „quaritur de loco Angeli“ einleitend in drei Teilfragen aufgespalten; mit einer solchen, allerdings ganz anders motivierten Aufspaltung fängt auch unsere Erläuterung des Grundgedankens „Ein wohlgeordneter Text beantwortet eine Quaestio“ an (S. 164):

## Die drei Fragen

- (1) Was geschah dann?
- (2) Wohin legte Lotte ihre Hand?
- (3) Wer legte die Hand auf die meinige?

stellen den Gefragten vor drei unterschiedliche Alternativen und verlangen von ihm, die jeweilige Alternative zu entscheiden. Unter „Alternative“ verstehen wir dabei die Wahl zwischen zwei oder mehr Möglichkeiten. Im ersten Fall ist der Gefragte aufgefordert, eines von verschiedenen denkbaren Geschehnissen anzugeben, die zu einer bestimmten Gelegenheit hätten passiert sein können — wobei diese Gelegenheit selbst nur ungefähr angedeutet wird („dann“). Im zweiten soll er einen unter den verschiedenen Orten angeben, auf die Lotte bei jener (nur aus dem Kontext erschließbaren) Gelegenheit ihre Hand gelegt haben könnte. Und im dritten soll er angeben, welche Person es war, die bei einer bestimmten gleichfalls nur dem Kontext entnehmbaren Gelegenheit die Hand auf die des Fragers gelegt hat. Alle drei Alternativen kann der Gefragte dadurch entscheiden, daß er den Satz

- (4) Lotte legte ihre Hand auf die meinige.

äußert. Allerdings ist in diesen drei Fällen die Intonation deutlich verschieden. Dies spiegelt eine Aufgliederung der gesamten Äußerung in zwei unterschiedliche Komponenten wider — eine erste, die wiederum die Alternative angibt, die es zu entscheiden gilt, und eine zweite, die angibt, welche der Alternanten der Sprecher tatsächlich wählt, d. h. wie er die anstehende Alternative spezifiziert. Eine solche Alternative, die zur Entscheidung ansteht, bezeichnen wir als die Topik einer Äußerung, jene Alternante, die dann tatsächlich gewählt wird, als ihren Fokus. Im letzten Fall, also nach (2), ist der Fokus jener Ort, der in (4) durch den Ausdruck „auf die meinige“ ungefähr bezeichnet wird, im dritten ist es jene Person, auf die in (4) durch den Ausdruck „Lotte“ Bezug genommen wird, im ersten schließlich das Geschehnis, welches der gesamte Satz (abzüglich des auf die Vergangenheit verweisenden Tempusmorphems) zum Ausdruck bringt. Die Begriffe Topik und Fokus, wie wir sie hier verwenden, beziehen sich auf Bedeutungen, nicht auf Ausdrücke. Man muß daher zwischen Topik und Fokus einerseits, Topikfestlegung und Fokusfestlegung andererseits unterscheiden.

Diese Definition von Topik und Fokus entspricht keiner der damals und oftmals auch heute noch gängigen Vorstellungen, also beispielsweise der auf Georg von der Gabelentz zurückgehenden Idee, dass es etwas gibt, worüber

etwas gesagt wird, und etwas, das darüber gesagt wird („aboutness“), oder der auf Henri Weil zurückgehenden Idee, dass die Rede von etwas Bekanntem zu etwas Unbekanntem fortschreitet („given – new“); beides findet sich in vielen Varianten und Kombinationen. Die Idee hier ist vielmehr, dass eine zu entscheidende Alternative aufgemacht wird, aus der dann ein Element ausgewählt wird; die Alternative kann dabei mehr als zwei Elemente enthalten. Im einfachsten Fall entspricht das der Relation Frage : Antwort. In Wirklichkeit ist es weitaus komplexer, denn diese Dichotomie kann sich ganz oder teilweise innerhalb eines Satzes finden. In unserem Aufsatz geht es nicht vorrangig um Bedeutung und Form von Topik und Fokus; die Dichotomie wird dort primär dafür genutzt, um die referentielle Bewegung in der Hauptstruktur von Erzählungen – den „narrative thread“ – zu erfassen.

Seither sind fast vier Jahrzehnte vergangen, es ist viel, sehr sehr viel zur Informationsstruktur geforscht worden, und man würde hoffen, dass nun eine gewisse Klarheit bei Fakten und Begriffen eingekehrt ist. Aber es verhält sich wohl eher so, wie einmal ein kenntnisreicher Literaturwissenschaftler über den Ursprung des Amadis-Romans – des Musters aller Ritterromane – gesagt hat: Je mehr man in dieser Sache geforscht hat, umso unklarer wurde sie. Das umfassende Handbuch von Fery und Ishihara (2016) vermittelt ein beeindruckendes Bild von dem bestehenden Wirrsal an Beobachtungen und Begriffen. Im Folgenden möchte ich daher nicht versuchen, Begriffe wie Topik, Kommentar, Thema, Rhema oder Fokus einer neuen Sektion zu unterziehen, sondern einen Weg gehen, bei dem solche Begriffe sich eher als Spezialfälle eines allgemeineren Phänomens erweisen. Der Ansatz dazu ist die Idee der Alternative, aus der ein Element ausgewählt wird.

Nach dieser Idee wird, wenn Gedanken in Worte umgesetzt werden, fortwährend eine Auswahl aus verschiedenen Alternativen getroffen. Dazu ist dreierlei erforderlich:

- die „Auswahlmenge“ muss klar sein;
- das „ausgewählte Element“ muss klar sein;
- es muss klar sein, *dass* dies das ausgewählte Element ist.

Wie immer in der menschlichen Kommunikation können die erforderlichen Informationen aus dem lexikalischen Gehalt der einfachen oder zusammengesetzten Ausdrücke stammen oder aber, nicht weniger wichtig, aus dem Kontext, also aus dem, was zuvor oder danach gesagt wird, aus dem Situationswissen oder aus dem allgemeinen Weltwissen der Beteiligten.

Das deutsche Wort „Alternative“ wird leider in zwei Weisen verwendet: Einmal meint man damit eine Menge von zwei oder mehr einander ausschließenden Möglichkeiten, von denen eine ausgewählt werden kann (*Wir stehen nun vor der folgenden Alternative*), oder man meint eine dieser Möglichkeiten

(Zu diesem Verbot gab es seinerzeit keine Alternative). Im Folgenden wird wie schon im Aufsatz von 1987 „Alternative“ für die Auswahlmenge verwendet und „Alternanten“ für die verschiedenen Elemente, aus denen sie besteht. Wesentlich ist nun, dass man bei sprachlichen Ausdrücken mindestens drei Arten von Alternativen und entsprechend ihren Alternanten unterscheiden muss. Das kann man sich an dem Ausdruck *drei grüne Blätter vor Augen führen* (ich versuche, das so zu beschreiben, dass es trivial erscheint – ist es aber nicht):

Typ A. Die meisten lexikalischen Einheiten haben mehrere Lesarten. Man hat bei *drei grüne Blätter* also eine Alternative wie (mindestens) {*blatt*-Laub, *blatt*-Papier, *blatt*-Säge} ebenso wie eine Alternative wie (mindestens) {*grün*-Farbe, *grün*-Erfahrung, *grün*-Politik}; keine Lesarten und mithin keine Alternative gibt es bei *drei*. Solche „Lesartenalternativen“ bestehen also innerhalb einer bestimmten Wortform, es sind „intralexikalische Alternativen“.

Typ B. Die zweite Art von Alternativen besteht darin, dass ein Wort (und analog ein zusammengesetzter Ausdruck) immer im Gegensatz zu Wörtern steht, die eine andere phonetische oder graphematische Form haben. Das ist die klassische Idee der strukturellen Semantik: {*blatt*, *zweig*, *knospe*, *blüte*, ...} bzw. {*grün*, *rot*, *farblos*, *bunt*, ...}. Hier sind die Alternativen also „interlexikalisch“. Genau besehen ist dies eine Vereinfachung, denn der Gegensatz besteht normalerweise nicht zwischen verschiedenen lexikalischen Einheiten, sondern zwischen bestimmten Lesarten verschiedener lexikalischer Einheiten. Das ändert aber nichts an der grundsätzlichen Unterscheidung.

Typ C. Die dritte Art von Alternativen rührt daher, dass verschiedene Entitäten durch ein und denselben Ausdruck beschrieben werden können, also die von ihm angegebenen Eigenschaften haben. Es gibt schon bei einem einzigen Baum, der 300 000 Blätter haben mag, Milliarden von Entitäten, die unter die Beschreibung *drei grüne Blätter* fallen: {*grünes blatt*-1953, 1987, 18951, *grünes blatt*-34, 4711, 69117, *grünes blatt*-206421, 4, 33, ...}. Es sind dies also „referentielle Alternativen“ oder, wie ich im Folgenden sagen will, „extralexikalische Alternativen“. In diesem Beispiel sind die Alternanten Objekte, also von jener Art, die man gewöhnlich durch Nominalphrasen beschreibt. Es kann sich aber auch um Ereignisse, Prozesse, Aktivitäten, Zustände, kurzum um Situationen handeln; es gibt Milliarden von Situationen, die unter die Beschreibung *Es war kalt*. fallen, weil es endlos viele Zeiten und Orte gibt, an denen es kalt war.

Man muss also zwischen intralexikalischen, interlexikalischen und extralexikalischen Alternativen unterscheiden. In allen drei Fällen erfolgt die

Auswahl durch Informationen aus dem Kontext. Dazu zählt natürlich auch der unmittelbare sprachliche Kontext, insbesondere der im selben Satz wie in *die drei grünen Blätter, die da auf dem Tisch liegen* oder *Am 14. Oktober 1953 war es in München kalt*. Die sprachlichen, situativen und aus dem Weltwissen rührenden Informationen bestimmen aber im konkreten Fall nicht nur die Auswahl aus einer gegebenen Alternative, sondern auch, was alles zu dieser Alternative gehört. In einer bestimmten Redesituation sind interlexikalische Alternanten zu *grün*, die nicht gleichfalls eine Farbe sind, völlig irrelevant und daher aus der Auswahlmenge ausgeschlossen, und ebenso sind in einer bestimmten Situation Zeiten und Orte aus der Auswahlmenge ausgeschlossen, in denen es zwar auch potenziell kalt war, die aber im Redezusammenhang keine Rolle spielen. All dies verlangt nach vielen weiteren Erläuterungen, aber es reicht für das, was ich zeigen will, nämlich welche Rolle dieser Mechanismus von Auswahlmenge, ausgewähltem Element und der Angabe seines Status in der Informationsstruktur des einfachen Satzes („clause“) und bei der Definitheit von Nominalphrasen spielt.

Mein erstes Beispiel geht auf eine Einsicht zurück, zu der ich vor etlichen Jahren gekommen bin (siehe z. B. Klein, 2006), allerdings ohne jeden Bezug zu der Idee der Auswahl aus Alternativen. Ist die folgende Assertion wahr oder falsch?

- (1) Kein Geistlicher hat ihn begleitet.

Auf diese Frage gibt es keine Antwort, solange man nicht weiß, von welcher Situation die Rede ist. Wenn die Situation, um die es geht, die Beerdigung von Papst Benedikt XVI. ist, dann ist die richtige Antwort „falsch“ (nachgerade grottenfalsch). Geht es hingegen um die Beerdigung von Werther, dann ist die richtige Antwort „wahr“ (jedenfalls wenn die ganze Geschichte wahr ist). Daraus folgt etwas ganz Wesentliches:

*Jede Assertion ist relativ zu der Situation, über die etwas gesagt wird; daher ist es nicht sinnvoll zu sagen, dass eine Assertion wahr oder falsch ist, wenn nicht klar ist, über welche Situation gesprochen wird.*

Eigentlich ganz offenkundig. Es gilt für Hauptsätze wie *Kein Geistlicher hat ihn begleitet*. Es gilt analog auch für Nebensätze wie *Wenn ihn kein Geistlicher begleitet hat*, mit dem wichtigen Unterschied, dass in diesem Fall explizit als **unentschieden** markiert wird, ob die betreffende Situation diese Eigenschaften hat oder nicht; in einer Assertion hingegen wird markiert (typischerweise durch die Position des finiten Elements und durch den finalen Intonationsfall), dass die Situation, über die geredet wird, die betreffenden Eigenschaften hat.

Hier ein etwas anderer Fall, der den entscheidenden Punkt verdeutlicht. Eines frühen Morgens schaut Henriette aus dem Fenster und sagt zu Margarete:

(2) Es schneit.

Diese Aussage bezieht sich also auf die Landschaft, die die Sprecherin gerade sieht; eine entsprechende Frage hätte sein können: „Was siehst du gerade draußen?“ Das ist hinsichtlich der Situation, über die gesprochen wird, etwas unscharf, z. B. wie weit das gerade Gesehene reicht. Dies ist typisch für natürliche Sprache – es gibt einen gewissen Interpretationsspielraum, der oft, aber nicht immer, durch Kontextinformationen gefüllt wird. In jedem Fall bezieht sich die Behauptung auf eine bestimmte Situation, und ihr Wahrheitsgehalt muss relativ zu dieser Situation beurteilt werden. In einer Sprachsituation wie der vorliegenden ist es trotz der verbleibenden Unschärfe sicherlich klar genug, über welche Situation geredet wird.

Wie kann man wissen, über welche Situation geredet wird? Die notwendigen Informationen können sich aus dem Kontext ergeben, z. B. durch eine vorangehende Frage („Was wissen Sie über *Werthers Leiden*?“), durch die Tatsache, dass die Äußerung Teil eines längeren Textes ist (z. B. auf den Satz *Am 31. Mai wurde er begraben* folgt), oder durch die kommunikative Situation, z. B. wenn jemand bei einem Fußballspiel zu seinem Nachbarn *Fantastisch!* über die Aktion des Torhüters sagt, die sie gerade gesehen haben. Oder sie wird von der Äußerung selbst geliefert; in diesem Fall wird ein Teil ihres beschreibenden Inhalts verwendet, um die Situation, über die geredet wird, zu identifizieren. Der andere Teil wird genutzt, um die Informationen anzugeben, die der so identifizierten Situation zusätzlich zukommen. Dies gilt für Assertionen; bei der entsprechenden Frage, hier also „Hat ihn ein Geistlicher begleitet?“ hat man dieselbe Aufteilung in einen Identifikationsteil und einen Zusatzteil, und es wird als nicht entschieden, sondern als zu entscheiden markiert, ob ersterer letzterem zukommt; beispielsweise durch ein Fragewort, durch Voranstellung des Finitums oder durch die Intonation. Und ebenso kann markiert werden, dass dies unentschieden ist, z. B. in einem „wenn-Satz“. In jedem Satz hat man mithin

- einen „I-Teil“, der die Situation, über die geredet wird, festlegt, soweit sich das nicht bereits aus dem Kontext – etwa einer expliziten Frage – ergibt; der I-Teil liefert großenteils gegebene Information, die daher beibehalten werden kann und typischerweise auch wird;
- einen „Z-Teil“, der weitere mögliche Eigenschaften angibt; dabei handelt es sich typischerweise um „neue“ Information, die das, was über die Situation bereits bekannt ist, bereichert;
- einen „R-Teil“, der angibt, wie die Relation zwischen beiden ist.



Ich habe das hier ohne Bezug auf die traditionelle Terminologie dargestellt; aber es liegt nahe, die Situation, über die geredet wird, als die „Topiksituation“ zu bezeichnen und den Z-Teil als den „Kommentar“ (nach manchen auch als Fokus oder als Rhema); für den R-Teil gibt es kein entsprechendes Gegenstück. Aber Begriffe wie „gegeben / alt / bekannt / vorausgesetzt“ versus „neu / unbekannt“ sind hier keine Definitionsmerkmale, sondern Eigenschaften, die typischerweise mit der Funktion von I-Teil und Z-Teil einhergehen, aber nicht müssen. Auch ist hier die „Topik“ nicht ein Teil des Satzes, beispielsweise das Subjekt; vielmehr ist es eine ganze Situation, deren bisher angegebene Eigenschaften nun noch ergänzt werden.

Was hat all dies nun mit der Idee der Alternative und der zu wählenden Alternante zu tun? Es ist nichts als ein spezieller Fall davon. Der I-Teil legt eine Menge von Alternanten fest, nämlich die möglichen Z-Teile; der R-Teil markiert dann den tatsächlich angegebenen Z-Teil als ausgewählt, bei einer Assertion also als „hat diese Eigenschaften“. Man kann sich das ein wenig wie ein Porträt vorstellen, das schon weithin gemalt ist; aber eine Stelle ist freigelassen, um wahlweise einen Borsalino, einen Bowler, eine Mitra oder einen Steyrerhut einzufügen; eines davon wird ausgewählt, reingemalt und dadurch als ausgewählt markiert.

Die Alternative ist in diesem Fall von Typ B – es geht um einen Gegensatz zwischen lexikalischen Alternanten, die der Situation zugeschrieben werden können. Bei einer W-Frage wie *Was lag unter dem Baum?* könnte sie – je nach weiterem Kontext – beispielsweise {ein grünes Blatt, drei grüne Blätter, vier rote Blätter, sieben tote Mäuse, ein Gartenzwerg, ...} sein. Dasselbe gilt typischerweise für einen Deklarativsatz *Daneben sah man ...*<sup>3</sup> Bei einer Interrogativfrage wie *War es kalt?* ist – sehr vereinfacht gesagt – die Alternative {*kalt sein*, *diff-kalt sein*}, wobei mit „diff-kalt“ alle von „kalt“ verschiedenen, aber kontextuell plausiblen Eigenschaften gemeint sind; hier gibt es mancherlei Komplikationen, auf die ich hier nicht eingehe. Die Alternative ist also gleichfalls interlexikalisch.<sup>4</sup>

In meinem zweiten Beispiel für den Alternative-Alternante-Mechanismus geht es hingegen um extralinguistische Alternativen (Typ C), also um die Wahl zwischen Entitäten, die unter die **gleiche** Beschreibung fallen. Das können

---

3 *Ceterum censeo*: Es spielen immer auch Informationen aus dem weiteren Kontext mit.

4 Am Rande sei hier angemerkt, dass in der „Alternative Semantics“, wie sie von Mats Rooth, Arnim von Stechow, Manfred Krifka und anderen entwickelt wurde, die Alternativen interlexikalisch sind. Das ist überhaupt die Voraussetzung dafür, dass der Fokus prosodisch hervorgehoben ist. Wenn die Alternative aus Alternanten besteht, die allesamt durch den gleichen Ausdruck beschrieben werden, kann man den Gegensatz zwischen ihnen nicht prosodisch hervorheben. Man kann sagen *Er hat keinen Hund gekauft, sondern ein Pferd*, aber es hat keinen Sinn zu sagen *Er hat kein Pferd(-Taylor) gekauft, sondern ein Pferd(-Rose)*.

Situationen sein, es gibt, wie schon bemerkt, Milliarden von Situationen, die durch *Es war kalt*. korrekt beschrieben sind. Im Folgenden geht es aber nur um unbelebte oder belebte Objekte, also Gegenstände oder Lebewesen, die gewöhnlich durch Nominalphrasen beschrieben werden.

Eine solche Alternative ist beispielsweise {*ein Pferd-Taylor, ein Pferd-Rose, ein Pferd-Rosinante*}, eine andere Alternative ist {*zwei Pferde-Rose-Taylor, zwei Pferde-Rosinante-Halla*} und wieder eine andere {*Pferde-jede Menge, die aus mehr als einem Pferd besteht*}<sup>5</sup>. Die Ausdrücke *ein Pferd, zwei Pferde, Pferde* sind allesamt unbestimmt in Bezug darauf, welche Entitäten nun tatsächlich gemeint sind: Sie sind in dieser Hinsicht „indefinit“. Das heißt ganz einfach, dass die durch den Ausdruck (und den Kontext!) spezifizierte Alternative verschiedene Alternanten enthält; er ist referentiell mehrdeutig. Um diese Mehrdeutigkeit aufzulösen, müssen Informationen aus dem Kontext herangezogen werden. Das ist beispielsweise der Fall, wenn der indefinite Ausdruck im Skopus eines anderen Ausdrucks steht, der eine Alternante festlegt, beispielsweise *Auf der Weide standen zwei Pferde*. Festgelegt wird sie, weil der Satz nur eine einzige Situation – zu einer bestimmten, im Satz aber nicht genannten Zeit in der Vergangenheit auf der Weide stehen – beschreibt, und relativ zu dieser Situation geht es um zwei bestimmte Pferde, nämlich die beiden, die da auf der Wiese standen. Deutlicher kann man dies machen, wenn man die Singularität der Situation durch ein entsprechendes Adverbiale explizit macht, wie etwa *Als ich ankam, standen zwei Pferde auf der Weide*. Hingegen ist mit *Dreimal standen, als ich ankam, zwei Pferde auf der Weide* nichts darüber gesagt, ob es dreimal dieselben sind oder nicht. Allerdings macht dieser Satz eine Auswahl für jede einzelne dieser drei Situationen; es gibt also wohl eine Singularität, aber sie bezieht sich auf die jeweilige Situation.

„Indefinit“ heißt bei dieser Betrachtungsweise einfach „der Ausdruck definiert eine Alternative, wählt aber keines ihrer Elemente aus“. Traditionell stellt man indefiniten Ausdrücken definite gegenüber, oft in Verbindung mit der Unterscheidung zwischen indefinitem und definitem Artikel: *ein Pferd – das Pferd*. Das ist allerdings irreführend, denn man hat ja denselben Gegensatz bei *zwei Pferde – die zwei Pferde, viele Pferde – die vielen Pferde*, und ganz einfach *Pferde – die Pferde*. Der „unbestimmte“ Artikel ist lediglich ein Zahlwort, das beim Singular ausgelassen wird, es sei denn, man verwendet es kontrastiv: *das eine Pferd – das andere Pferd*. Was aber leistet der „bestimmte“ Artikel in all diesen Fällen? Darüber gibt es eine reiche Literatur und eine Reihe von Theorien, zu denen allesamt es viele nicht-triviale Gegenbeispiele gibt (vgl. dazu Klein, 2000). Die Idee der Alternativenauswahl gibt eine recht einfache Antwort: Die Nominalphrase ohne definiten Artikel wie *zwei Pferde* gibt eine Alternative an; der Hinzufügung eines definiten Artikels wie in *die*

---

5 *Ceterum censeo*: Der Kontext liefert gleichfalls Informationen.

*beiden Pferde* markiert, dass eine Auswahl daraus getroffen ist. Der Artikel sagt allerdings nicht, was die ausgewählte Alternante ist, kann er auch gar nicht, denn alle Alternanten werden gleich beschrieben: Der Ausdruck *die zwei Pferde* hat keinen größeren deskriptiven Gehalt als der Ausdruck *zwei Pferde* allein, ebenso hat *die vielen Pferde* keinen höheren deskriptiven Gehalt als *viele Pferde* allein. Man weiß aber immerhin, dass eine Auswahl getroffen ist und der Ausdruck relativ zur Alternative referentiell singular sind. Will man wissen, was die Alternanten nun sind, muss man wiederum jenseits des Ausdrucks selbst zu Rate ziehen. Es mag allerdings sein, dass die Alternative selbst nur ein Element enthält, wie beim Superlativ: *der höchste Berg in Europa*. Es gibt nur eine Entität, die durch *höchster Berg in Europa* richtig beschrieben werden kann, und deshalb kann man da die Markierung „Auswahl gemacht“ setzen. Wohl gibt es auch eine Gruppe *drei höchste Berge in Europa*, und diese Gruppe (!) ist wiederum singular; deshalb kann man sagen *die drei höchsten Berge in Europa*, obwohl nicht jeder Berg am höchsten ist.

Wenn Ausdrücke mit dem definiten Artikel wie *das braune Pferd* in der traditionellen Literatur als „einzig“ oder als „beibehalten“ charakterisiert werden, dann liegt dies einfach daran, dass *braune Pferd* nur eine Alternative angibt, während *das braune Pferd* zusätzlich angibt, dass daraus ein Element ausgewählt ist. Es sagt nicht, was dieses Element ist, es sagt nur „such im Kontext!“ oder „schau, was die Situation ist!“. Für die Singularität haben wir das eben gesehen. Ein typisches Beispiel für die „beibehalten“-Vorstellung sind Folgen wie *Auf der Weide standen ein braunes und ein schwarzes Pferd. Das braune Pferd ...* Dass man das braune Pferd beim zweiten Vorkommen als „beibehalten“ (oder „anaphorisch“) versteht, liegt nicht daran, dass der definite Artikel Anaphorizität ausdrückt, sondern daran, dass er markiert „ein braunes Pferd aus der Alternative ausgewählt“, und das einzige Pferd, das dafür in Frage kommt, ist das im vorigen Satz als auf der Weide stehend genannte. Die vieldiskutierten Eigenschaften, die mit der Definitheitsmarkierung einhergehen, sind daher für diese nicht definitorisch, es sind natürliche Konsequenzen des Spiels von Alternative und Auswahl aus der Alternative.<sup>6</sup>

---

6 Im Jahre 1905 entwickelte Bertrand Russell seine berühmte Singularity-Theory des definiten Artikels. Er beschließt seinen Essay mit den Worten: „Of the many other consequences of the view I have been advocating, I will say nothing. I will only beg the reader not to make up his mind against the view – as he might be tempted to do, on account of its apparently excessive complication – until he has attempted to construct a theory of his own on the subject of denotation. This attempt, I believe, will convince him that, whatever the true theory may be, it cannot have such a simplicity as one might have expected beforehand.“

Es ist schon erstaunlich, dass es auch heute, mehr als ein Jahrhundert danach, für *the* – das mit Abstand häufigste Wort der mit Abstand meistuntersuchten Sprache der Welt –, keine allgemein akzeptierte Analyse gibt, schon gar keine einfache. Hier ist

Damit kehre ich noch einmal zurück zum Ausgangspunkt, nämlich der Art, wie die „Topik-Fokus-Struktur“ in Klein und von Stutterheim (1987) beschrieben wurde:

Dies spiegelt eine Aufgliederung der gesamten Äußerung in zwei unterschiedliche Komponenten wider – eine erste, die [...] die Alternative angibt, die es zu entscheiden gilt, und eine zweite, die angibt, welche der Alternanten der Sprecher tatsächlich wählt, d. h. wie er die anstehende Alternative spezifiziert.

Dort ging es nur um die Frage, wie die Quaestio eines Textes dessen Struktur vorgibt. Was dort angedeutet wurde, ist aber weitaus fruchtbarer; es kann eine Reihe von vieldiskutierten Problemen der Sprachtheorie auf recht einfache Weise lösen, oder zumindest einen Weg dahin zeigen.

## 7 Excipit

Nicht jeder liebt Stifters *Nachsommer*. Es ist ein Buch, in dem auf vielen hundert Seiten wunderbarer Prosa wenig passiert, außer dass ein junger Mann herausfindet, wie er leben soll. Er sucht es nicht bewusst, es ergibt sich gleichsam von selbst. Am Ende heißt es dann:

Was mich selber anbelangt, so hatte ich [...] die Frage an mich gestellt, ob ein Umgang mit lieben Freunden, ob die Kunst, die Dichtung, die Wissenschaft das Leben umschreibe und vollende oder ob es noch ein Ferneres gäbe, das es umschließe und es mit weit größerem Glück erfülle. [...] Ob ich es nun in der Wissenschaft, der ich nie abtrünnig werden wollte, weit werde bringen können, ob mir Gott die Gnade geben wird, unter den Großen derselben zu sein, das weiß ich nicht; aber eines ist gewiß, das reine Familienleben [...] ist gegründet, es wird, wie unsre Neigung und unsre Herzen verbürgen, in ungeminderter Fülle dauern, ich werde meine Habe verwalten, werde sonst noch nützen, und jedes, selbst das wissenschaftliche Bestreben, hat nun Einfachheit, Halt und Bedeutung.

So schön denkt sich einer eine Antwort auf die Quaestio *Wie sollen wir leben?* aus. Aber das Leben hat seine eigenen Antworten.

---

ein Vorschlag: Das Wort *the* in *the xyz* markiert, dass aus der von *xyz* beschriebenen Auswahlmenge genau ein Element ausgewählt ist. Es sagt nicht, welches; dazu muss man den Kontext zu Rate ziehen.

## Literatur

- Babusiaux, U. (2011). *Papinians Quaestiones: Zur rhetorischen Methode eines spät-klassischen Juristen*. C. H. Beck. <https://books.openedition.org/chbeck/1212>.
- Fery, C. & Ishihara, S. (Hrsg.) (2016). *The Oxford Handbook of Information Structure*. Oxford University Press.
- Hoye, W. J. (1997). Die mittelalterliche Methode der Quaestio. In N. Herold, B. Kensmann & S. Mischer (Hrsg.), *Philosophie: Studium, Text und Argument* (S. 155–178). Lit.
- Klein, W. (2000). Was uns die Sprache des Rechts über die Sprache sagt. *Zeitschrift für Literaturwissenschaft und Linguistik*, 118, 115–149.
- Klein, W. (2006). On finiteness. In V. Van Geenhoven (Hrsg.), *Semantics in acquisition* (S. 245–272). Springer.
- Klein, W. & von Stutterheim, C. (1985). *Text structure and referential movement*. MPI für Psycholinguistik Nijmegen.
- Klein, W. & von Stutterheim, C. (1987). Quaestio und referentielle Bewegung in Erzählungen. *Linguistische Berichte*, 108, 163–183.
- Miller, M. & Klein, W. (1981). Moral argumentations among children: A case study. *Linguistische Berichte*, 74, 1–19.
- von Stutterheim, C. (1986). *Temporalität in der Zweitsprache. Eine Untersuchung zum Erwerb des Deutschen durch türkische Gastarbeiter*. De Gruyter.
- von Stutterheim, C. (1997). *Einige Prinzipien des Textaufbaus. Empirische Untersuchungen zur Produktion mündlicher Texte*. Niemeyer.
- von Stutterheim, C. & Klein, W. (1986). A concept-oriented approach to second language studies. In C. Pfaff (Hrsg.), *First and second language acquisition processes* (S. 191–205). Cambridge University Press.
- von Stutterheim, C. & Klein, W. (1989). Referential movement in descriptive and narrative discourse. In R. Dietrich & C. F. Graumann (Hrsg.), *Language processing in social context* (S. 39–76). Elsevier.



---

Christine Dimroth, Cecilia Andorno, Sandra Benazzo

# When Discourse Elicitation Tasks Go Dialogue: Introducing Entities in French, German, and Italian

---

**Abstract** This paper presents first results from a study on oral picture description dialogues produced by native speakers of French, German, and Italian. The aim is to find out how speakers of these languages introduce the objects they want to talk about so that their interlocutors can identify and localise them while building a coherent representation of the picture. In doing so, we are tying in with work on the description of spatial configurations by Christiane von Stutterheim and colleagues (Carroll & von Stutterheim, 1993; von Stutterheim, 1997a), who used a similar method, keeping the stimulus and the procedure constant while running the task with speakers of different languages. When compared with the highly sophisticated non-verbal data elicitation techniques used by the Heidelberg team in their more recent work on event cognition, asking someone to describe a picture at their own pace seems quite a trivial approach. We are convinced, however, that it is worthwhile to go back to (some of) the roots, building on earlier work and at the same time changing a parameter that has quite some impact on the course of events: Instead of quasi-monologues with a rather passive addressee, we are studying dialogues in which two speakers are acting at eye level when describing and comparing mutually unknown spatial configurations. The reconstructed *Quaestio* (Klein & von Stutterheim, 1987) that is assumed to support a speaker's selection and organisation of information in a monologue might be locally overwritten in a dialogue when an interlocutor's real questions or statements alter the information flow.

**Keywords** Discourse, dialogue, typology, information structure, spatial description, entity introduction and contrast

## Introduction

Describing a picture so that a listener can mentally reconstruct it requires that the speaker produces subsequent discourse moves during which he or she identifies a specific object (the *Entity-Theme*) and indicates its position in relation to something the hearer can localise because it is already part of the common ground (the *Relatum*). Von Stutterheim & Carroll (1993) and Carroll & von Stutterheim (1997) identified different strategies speakers use to establish a spatial relation between a Theme and a Relatum when describing static configurations. When speakers follow a global strategy, they localise entities relative to salient regions of the overall scenery. In a picture description task, these are regions of the picture itself, e.g., its borders or corners. When speakers follow a local point-by-point strategy, the Relatum is the space occupied by an entity that both interlocutors can identify. In addition, speakers can adopt a linear strategy and describe the content as if they were following an imaginary path. In all cases, speakers will also take care to signal the information structure of the utterance and the relation of its elements to the preceding discourse. In the introduction of a Theme, for example, the relevant entity will, per definition, be new to the discourse, and the expression referring to it will constitute the focus of the utterance as in (1a). The expression encoding the Relatum, on the other hand, might also be new, but in a coherent description, it will probably bear some anaphoric relation to what was said before. What is used as the Relatum might, for example, be an entity that represented the Theme or the Relatum of an earlier utterance, as in (1b). In picture description monologues, the expression referring to the Theme and the expression referring to a position identified with the help of the Relatum are the two central information units. We will call them Entity and Localisation, respectively.

(1a) There is a teapot [Entity] in the top right corner [Localisation]

(1b) Below the teapot [Localisation] there is an orange [Entity]

In the current study, two speakers interact while describing two (slightly) different pictures to each other in a “spot-the-difference-task”. The pictures show random collections of everyday objects that are partly similar and in the same location, partly similar and in a different location, and partly dissimilar but in the same location (see Appendix). Each participant could only see his or her own picture, and both were instructed to jointly detect as many differences between their pictures as possible.

We assume that this manipulation has consequences on three different levels: (1) a **collaborative discourse construction**, (2) an **extra information unit**, and (3) the possibility of **contrastive content**.



A **collaborative discourse construction** involves joint common ground management (Stalnaker, 2002). Both speakers are contributing content and they must react to the interlocutors' discourse moves, e.g., answer explicit and implicit questions concerning the existence and the location of entities or express confirmation<sup>1</sup> and disconfirmation of descriptions proposed by the interlocutor, until they get to a shared representation of similarities and differences.

A disconfirmation does, however, not mean that speakers disagree with each other because there are two different pictures. Diverging statements are thus not mutually exclusive but will eventually become part of a shared situation model involving a representation of both pictures. When comparing their pictures, speakers might feel the need to indicate which picture they are talking about. Next to the Entity and its Location, there is thus **an additional information unit** to deal with. We will call this information unit the Frame of Reference<sup>2</sup> and consider all sorts of linguistic expressions speakers use to indicate the picture they are referring to (e.g., *in my picture ...; as for me, I have ...*).

Even though the presence of two Frames of Reference excludes disagreement, speakers must cope with signalling **contrastive content**, for example, in expressing that the situation in their own picture is partly different from a previous description produced by the interlocutor (*where you have an apple, I have a pear*). In the contexts we study, the information units Frame of Reference and Entity can be contrastive, whereas the Location talked about is not.<sup>3</sup> The resulting discourse relations might lead to an increase in marked structures. Overall, we assume that all three differences (co-constructing the discourse, referring to an additional information unit, and signalling contrasts) induce communicative effort.

As Christiane von Stutterheim and colleagues have shown, conceptual structure is not a simple reproduction of the speaker's memory representation of a spatial configuration, but perspective-driven and influenced by the speakers' communicative intention. A successful spatial description not only requires that the discourse moves contain the relevant information units but that they are tied together to form a coherent whole (von Stutterheim & Carroll, 1993; von Stutterheim, 1997b). When trying to fulfil these parallel requirements that have an impact on the selection and linearisation

---

1 See Dimroth & Starren (2022) for a study on confirmation based on data from the same task.

2 The Frame of Reference contributes to the identification of the two different "Topic Situations" (Klein, 2008) that are at stake in our task.

3 Utterances in which the Entity is maintained and the Location is contrasted occur in the data but are not considered here.

of information units, speakers rely on quasi-automatised routines. Cross-linguistic comparisons reveal that these routines are shaped by the lexical and grammatical resources available in a language. These resources provide a scaffolding for the conceptual material and lead the speaker to adopt the perspective that is most easily expressed. “What is striking here is that speakers so rarely make use of options that differ from the norm.” (Slobin, 1991, p. 17, cited in Carroll & von Stutterheim, 1993, p. 1012).

In our study, we will investigate how speakers use the different options at their disposal when reacting to their interlocutors’ contributions in a dialogue. The study of speakers adapting to listeners when presenting complex information also has a predecessor in Christiane von Stutterheim’s work. Von Stutterheim and Kohlmann (1998) conducted experiments in which speakers had to instruct listeners how to construct a toy village. During the experiment, the listener first shared the speaker’s perspective and later changed place (looked at the scene from with a deviation of 90°). Interestingly, speakers barely reacted to this change of position (see also the study by Speck, 1995). If the listener signalled comprehension difficulties, speakers tried to make their own viewpoint more clearly accessible to the listener but did not take the listeners’ perspective. These findings were interpreted as supporting a view according to which the macro-structural organisation of discourse moves is determined by the *Quaestio* and not accommodated to the listener’s perspective: “Das bedeutet, daß auf makrostruktureller Ebene angesiedelte ‘Entscheidungen’ des Sprechers hinsichtlich Kohärenz, Perspektive, Linearisierung, etc. keine Kandidaten für hörerbefugene Revisionsprozesse sind.” (von Stutterheim, 2001, p. 478).

Importantly, though, the listener in these experiments was a passive addressee (mostly a confederate speaker) who sometimes signalled misunderstandings but did not otherwise contribute actively to the success of the exchange. In the empirical study we present here, we investigate the interaction between two speakers who actively contribute to the communicative task. The participants alternate between the speaker and the listener role while both describe aspects of their own picture to the interlocutor in order to detect similarities and differences. Both speakers thereby adapt to their interlocutor’s contribution. In this sense, our “interlocutor adaptation” differs from the type of “hearer adaptation” investigated in these earlier studies. The main differences are the shared discourse construction, the extra information unit (Frame of Reference), and the presence of contrastive content.

## The languages in our study

Data were collected from native speakers of German, French, and Italian. The languages differ in many typological parameters concerning the basic sentence structure (see Table 1) that have an impact on the introduction of entities in different dialogue configurations.

Table 1. Basic grammatical properties in German, French, and Italian

	German	French	Italian
<i>word order</i>	V2	SVO	SVO
<i>word order flexibility</i>	+	-	+
<i>pro-drop</i>	-	-	+
<i>subj pronouns</i>	one series of pronouns for first and second person <sup>4</sup>	weak (clitic) and strong pronouns for all syntactic contexts	one series of overt pronouns for all syntactic contexts

The three languages also differ with respect to the constructions available for the locational predications that are at the centre of our interest in the current paper. Locational predications concern the spatial relation between a Theme (the entity whose position is described) and a Relatum (the entity used as a reference point for the spatial relation). Following Creissels (2019), we can identify two alternative locative predications differing in perspectivation: a Plain Locational Predication (PLP) selects the Theme (“Figure” in Creissels’ terminology) as its starting point (and possibly topic), whereas an Inverted Locational Predication (ILP) has the Theme as its endpoint and (part of) the focus<sup>5</sup>; see examples in Table 2. From a typological perspective, verb forms used for PLP and ILP constructions frequently align with other predicative types, such as:

- existential predications, concerning the existence or long-term presence of an entity in the world as in (2),
- equative predications, concerning an entity and its attributes as in (3), and

4 There are two series of pronouns (*er/der*) with slightly different functions in third person contexts, but these do not play a role in the current discourse configurations.

5 An alternative description in this much debated matter is provided by Koch (2012), who adopts the labels of Thematic Location (PLP) and Rhematic Location (ILP).

- possessive predications, concerning a relation between a possessor and a possessee as in (4).
- (2) There are many good books on Napoleon’s campaigns.
  - (3) The books on Napoleon’s campaigns are boring.
  - (4) Mary has a book on Napoleon’s campaigns.

Each of the languages considered in the present paper (German, French and Italian) has three verbal constructions for these predication types. They differ in their alignment with locational PLP and ILP predications, as shown in Table 2.

Table 2. Form and alignment of language-specific constructions for existential, locational, equative, and possessive predications

Type of Predication		Syntactic Construction		
		German	French	Italian
Existential predication		<i>it_give</i>	<i>there_have</i>	<i>there_be</i>
Locational predication	ILP predication	<i>be</i> <sup>6</sup>	<i>be</i>	<i>be</i>
	PLP predication			
Equative predication				
Possessive predication		<i>have</i>	<i>have</i>	<i>have</i>

As shown in Table 2, the three languages share the same syntactic construction for equative (*be*) and possessive predication (*have*) and they differ in the construction for existential predication (GER *it\_give*; FRE *there\_have*; ITA *there\_be*). There are also differences in the domain of locational predications, and these are most relevant for our study. None of the languages has a specialised form confined only to ILP and/or PLP and they differ in their “alignment”. In French and Italian, PLP aligns with the equative predication and ILP aligns with the existential predication; in German, they both align

6 German also has posture verbs that can be used in ILP and PLP constructions (e.g., *Der Baum steht vor der Kirche / Vor der Kirche steht ein Baum*). Their use is however not as systematic as it is in other Germanic languages (Czinglar 2002) and we will not discuss them in detail since they did not occur in our data. This is probably due to the display of the objects on our stimuli pictures that did not show any supporting surface (see below for details).

with the equative predication. As a consequence, only constituent order distinguishes ILP and PLP in German (see 5a and 5b).

- (5a) Das Buch von Marie **ist** auf dem Tisch. (PLP)  
 (5b) Auf dem Tisch **ist** das Buch von Marie. (ILP)

In French and Italian, a change in the lexical verb is required.<sup>7</sup>

- (6a) Le livre de Marie **est** sur la table (PLP)  
 (6b) Sur la table **il y a** le livre de Marie (\*Sur la table **est** le livre de Marie) (ILP)  
 (7a) Il libro di Maria **è** sul tavolo (PLP)  
 (7b) Sul tavolo **c'è** il libro di Maria (\*Sul tavolo **è** il libro di Maria) (ILP)

The aim of this study is to find out whether and how these cross-linguistic differences in the availability and the mapping of localising constructions onto predication types have an impact on speakers' preferred solutions for introducing entities in the dialogue task. Our study addresses the following research questions and hypotheses:

1. How do speakers introduce entities when they spontaneously mention them for the first time? We will refer to the relevant instances as First Move Introductions because they appear at the beginning of a sequence, i.e., they do not answer an overt preceding question (A: *What do you have at the bottom of the picture?* B: *A yellow teapot.*), and they are not in contrast to a preceding utterance about the same location (A: *There is an apple in the corner.* B: *In my picture, there is a pearl!*). We hypothesise that speakers of all three languages will mainly choose ILPs for First Move Introductions since these constructions allow them to start from information that is identifiable to the addressee before mentioning the new information. In addition to an adverbial identifying a position on the picture and the NP encoding the new (and focal) entity, speakers might find it relevant to indicate which picture they are talking about, i.e., to explicitly refer to one of the pictures or its owner (the Frame of Reference). The presence of this third information unit next to Entity and Location might lead to some competition for the topic role / position and it might also encourage a different choice concerning the locational predications adopted when compared to First Move Introductions in monologic descriptions.

---

<sup>7</sup> Note that constituent order can change in addition, i.e., a PLP could also be rendered as *Il y a le livre de Marie sur la table* / *C'è il libro di Maria sul tavolo*.

2. In Second Move Introductions speakers introduce entities in utterances that directly relate to a preceding utterance produced by their interlocutor. The preceding utterance can either be a question (*What do you have there?*) or it can be a First Move Introduction (*I have an orange to the right*). The question is, how Second Move Introductions differ from First Move Introductions. We assume that in replies to questions, Second Move Introductions are likely to be elliptical (*A teapot*). When following First Move Introductions, the Second Move can confirm or disconfirm the interlocutor's description. In the latter case, speakers often introduce a contrasting entity that occupies the position described by the interlocutor (*I have a tennis ball where your toothpaste is*). Contrastive Second Moves do not only assert a claim about the existence or the position of an entity, but also take a stance towards the similarity of the two pictures. They can, for example, contain an overt rejection of the content of the First Move Introduction (*No, I have an apple in the same place!*). We hypothesise that speakers adopt a different information structure when compared to First Move Introductions to signal the contrastiveness of (parts of) the information.

We assume that First and Second Move Introductions behave similarly with respect to some general properties of the three information units. The most central information unit in an introduction is the Entity (as we will see in the results, some introductions do not contain any other information). The Entity represents new information and is likely to be treated as focal (answering a question like *What is in your picture?*) in all cases. The information status of the other two information units (Location and Frame of Reference) could differ between First and Second Move Introductions. It is fair to say that the Frame of Reference (the picture talked about) can always be easily recovered from context: Even without any specific indication, it will be clear that speakers are talking about their own picture throughout the entire dialogue. It is, however, likely that speakers will more consistently mark the Frame of Reference in Second Move Introductions where it constitutes changing or contrastive information in addition to the Entity.

In contrast to the Frame of Reference, Location cannot be recovered on pragmatic grounds and is thus more informative in this sense. While the information about the location of an entity is new in First Move Introductions, it is maintained from the preceding context in Second Moves, where speakers talk about the same position (e.g., *the upper left corner*) on both pictures and contrast the entities encountered there.

In the following, we will compare First and Second Move Introductions in the three languages and describe whether and how they differ from introductions in monologues. The focus will be on the speakers' choice of predication

types / syntactic constructions and on the slots that these constructions open for the encoding of the three central information units (Entity, Location, Frame of Reference). In addition, we will look at the effect of information structure (topic, contrast, anaphoric relations) on the type of expressions that speakers chose and on their preferred word order in the given constructions.

## Empirical part: Method

Participants: For each language, 20 native speakers (university students, age range 19–32) were recruited. They were randomly assigned to one of ten dialogue pairs per language.

Procedure: The speakers in a dialogue pair were seated opposite of each other at a table. Each speaker received a picture that the other speaker could not see. The participants were informed that their pictures were not fully identical and that their task consisted of identifying as many differences as possible.<sup>8</sup> Speakers were not informed about the number of differences between the pictures. Note that there were no predefined speaker and addressee roles. Unlike comparable dialogue tasks (e.g., the “giver” and the “receiver” in Map Tasks; Anderson et al., 1991), participants had to establish a suitable division of labour on their own. They mainly took turns in providing their interlocutors with a description of some details of their pictures that were subsequently confirmed or disconfirmed by their interlocutors. Occasionally, this was explicitly negotiated in the beginning of the dialogue (*I start telling you what I see in my picture, OK?*), but roles typically shifted several times during the interaction.

Materials: The pictures (see Appendix) show photographs of twelve everyday household objects and food items (the “Entities”) that are arranged on a black background without any visible surface supporting them. Half of the twelve entities were similar, i.e., the same entities were in the same location on both photographs. Three entities were identical but were placed at a different position, whereas another three were at the same position on both pictures but differed in some property (e.g., full vs. empty toilet roll). Details are given in Table 3.

---

8 Whereas the German participants were told that they should complete the task within five minutes, there were no time constraints for the other groups. Given that most of the recordings in all three languages are between 4 and 5 minutes long, it is unlikely that this had a considerable impact on the descriptions.

Table 3. Individual entities depicted on the stimulus pictures according to comparative categories

Comparative category	N.	Depicted entities
same entity, same position	6	orange, tea pot, sellotape dispenser, plastic cup, tea light, cream jar
same entity, different position	3	tennis ball, toothpaste, pear
entity with different details, same position	3	empty vs. full toilet roll, real vs. wooden apple, yellow text marker vs. yellow crayon

Note that the pictures differ from materials used in earlier studies (description of a poster with a small town; Watorek, 2003; Carroll & von Stutterheim, 1997) in that a) the twelve entities are not located in any canonical spatial relation to each other, and b) the pictures are strictly 2-dimensional (there is a horizontal and a vertical axis, but no sagittal axis) and the objects do not have an intrinsic front-back or left-right orientation. Descriptions like “the tree is in front of the house” that rely on intrinsic object properties, or a depicted sagittal perspective are thus largely ruled out. The same holds for vague localisations like “next to the house there is a tree” where world-knowledge tells us that the most likely position of the Theme is left / right or in-front-of / behind, but not underneath or on top of the Relatum.

Carroll & von Stutterheim (1993) proposed to subdivide a Relatum’s space into an INNER space, an EXTERIOR space, a BOUNDARY space, and a NEIGHBOURING space. These categories do not apply to the possible Relata on our pictures where Themes are never located in or outside others or at their boundary. Instead, all items covered regions of roughly the same size (independently of their real size) and were presented in roughly equal distance to each other. In the absence of any logical or prototypical spatial relation between the entities, the only dimension that speakers could exploit was the NEIGHBOURING space that is not associated with any particular features of either Theme or Relatum.

### Database

With twelve entities and ten dialogues we expected 120 introductions per language (12 entities introduced in each dialogue). We found slight deviations from the expected numbers due to glitches occurring in the dialogues: speakers occasionally forgot that their interlocutor had already mentioned a particular entity and introduced it again; on other occasions, objects with



slightly different properties were treated as two entirely different entities and introduced independently.<sup>9</sup> It also happened that speakers introduced more than one item within one utterance (e.g., *To the right I have an apple and a teapot*). We counted these double introductions as one utterance if they met at least two of the following criteria that are supporting a list reading: they are marked by one continuous intonation contour; they are not interrupted by some back channeling signals from the interlocutor (*hmhm*); and are not linked to the preceding text by an overt connector. The resulting data is shown in Table 4.

Table 4. Data included in the analyses

	First Move Introductions	Second Move Introductions	Total
German	99	28	127
French	98	25	123
Italian	92	27	119

## Results Part I: First Move Introductions

In First Move Introductions, speakers spontaneously mention a particular item for the first time without contrasting the information they convey with something the interlocutor said before. Even though we are dealing with dialogue data, First Move Introductions are thus relatively independent of the interlocutor's contribution. The results can therefore be compared to findings from picture description monologues to isolate the impact of the additional information unit (Frame of Reference).

First Move Introductions often occur at the beginning of new discourse segments, typically after the interlocutors reached agreement about the position or properties of another entity. Utterances mentioning unspecified items (*There are twelve objects in my picture*) and utterances proposing a new label for a given entity (*You called this an orange, but I think it is a tangerine*) were not counted as Introductions.<sup>10</sup> In the following, we will only consider introductions in statements and exclude introductions in questions (*Do you*

9 Some speakers overlooked the differences between the two apples. Some others described the text marker and the crayon as two unrelated entities instead of two yellow pens with slightly different details.

10 See also note 11 concerning labelling and identification.

*also have an apple?*) due to their different syntax. The numbers of included utterances per language are given in Table 5.

Table 5. First Move Introduction: the subcorpus considered

	Total First Move Introductions	Introductions in questions	Introductions in assertions (considered)
<b>German</b>	99	1	<b>98</b>
<b>French</b>	98	11	<b>87</b>
<b>Italian</b>	92	11	<b>81</b>

When introducing entities, speakers can provide different amounts of details. They can simply assert that a particular entity exists on the picture they are describing (*There is a teapot*), or they can add information about the entities' location (*There is a teapot in the upper left corner*) and / or its properties (*I have a big yellow teapot*) in the same utterance. In our analyses, we largely neglect the description of properties and focus on the way entities are introduced and localised.

In our comparative analysis, we will first investigate language specific preferences for particular construction types in First Move Introductions (see the distinctions based on Creissels, 2019, that were proposed in the Introduction). The selection of a construction goes hand in hand with the availability of options for the expression of information units. In subsequent steps, we will report how often the information units Entity (ENT), Localisation (LOC) and Frame of Reference (FRA) are mentioned and where these information units are placed in terms of constituent order.

### Construction type

The following types of constructions occurred in First Move Introductions (see Table 2 in the Introduction):

<i>there_have</i> construction (FRE):	impersonal construction consisting of semantically bleached locative ( <i>y</i> ) + expletive subject ( <i>il</i> ) + verb of possessive predication <i>avoir</i> ; the two arguments encode the Entity (NP) and the Location (PP or adverbial);
<i>there_be</i> construction (ITA):	copular verb like in equative predication ( <i>essere</i> ) + semantically bleached locative ( <i>ci</i> ); the two arguments encode the Entity (NP) and the Location (PP or adverbial);

<i>be</i> construction:	copular verb (GER <i>sein</i> ; FRE <i>être</i> ; ITA <i>essere</i> ); the two arguments encode the Entity (NP) and the Location (PP or adverbial);
<i>have</i> construction:	transitive verb (GER <i>haben</i> ; FRE <i>avoir</i> ; ITA <i>avere</i> ) used in possessive predication; the two arguments encode the Entity (NP object) and the Frame of Reference (NP subject);
verbless construction:	information units are juxtaposed; Entity is encoded as NP; Location as adverbial or PP.

Concerning the possibilities to integrate the three information units, we can maintain that:

- LOC and ENT are largely similar across languages. LOC is always encoded as an adverbial / PP and ENT is always encoded as a nominal verb argument.
- The newly introduced ENT constitutes part of the focus. Across languages, PLP constructions (where ENT would be the topical starting point of the sentence) are therefore not well suited in these contexts.
- FRA is encoded as the first person subject argument of the verb in the *have*-construction. It can furthermore be encoded as an adverbial / PP (*in my picture*) in all constructions. The expressions differ in prominence: zero pronoun (Italian), obligatory clitic pronouns (French) and fully-fledged personal pronouns (German) are relatively light markers of FRA that come with the choice of a *have* construction. Stronger pronominal forms (overt pronouns in Italian or the strong pronoun *moi* in French) might have a similar “heaviness level” as explicit references to the speaker’s picture (*in my picture*). Note that forms can also be combined (*moi, sur mon image, j’ai une poire* ‘As for me, in my picture, I have a pear’).

Table 6 shows the frequency of the main construction types in the three languages. A small number of utterances (4 in German, 3 in French, 2 in Italian) contained individual lexical verb constructions (like *being located*, *being depicted*) that will be excluded from further analyses. Percentages (rounded) are given in brackets.

Table 6. Absolute numbers and frequency of construction types (First Move Introductions)

	<i>there_be/have</i>	<i>be</i> <sup>11</sup>	<i>Have</i>	<i>verbless</i>	<i>Total</i>
<b>German</b>	–	29 (31%)	15 (16%)	50 (53%)	94
<b>French</b>	15 (18%)	3 (3%)	62 <sup>12</sup> (74%)	4 (5%)	84
<b>Italian</b>	23 (29%)	1 (1%)	40 (51%)	15 (19%)	79

The first striking feature is the presence of *have* constructions and verbless constructions in all languages although in different proportions. As far as we can tell from the existing literature, there were no noteworthy amounts of these constructions in monological picture descriptions. Examples for the four main construction types attested in the dialogue data are given below.

*there\_be/have* construction:

- (8a) (FRE): juste au-dessus (**il**) y a un tube de dentifrice  
 (8b) (ITA): sotto c'è un barattolo blu

*be* construction:

- (9) (GER): rechts daneben **ist** bei mir so ein tesafilmabroller

11 Note that the three occurrences of the *be* construction in French and the one occurrence in Italian can actually be considered as borderline cases. The French utterances have a clitic subject (*ce*), e.g., *le dernier objet à droite c'est une bougie* ('the last object to the right is a candle'). We decided to count them as introducing (instead of a merely identifying) an entity, even though the indefinite NP is preceded by an unspecific reference ("object") in the left dislocation, that also hosts the localisation. The only occurrence of a *be*-construction in Italian is an even more doubtful case: The speaker starts the sentence as a (verbless?) locative predication, but then stops and wonders about the nature of the item she is introducing: *e sotto un # mmm suppongo sia un bicchiere delle macchinette del caffè* ('and below a # well I suppose it-is a cup for the coffee vending machines').

12 The total includes 4 occurrences of the lexical verb *voir* (see), which has the same argument structure as *avoir*.

*have* construction:

- (10a) (FRE): à droite du scotch j'**ai** une pomme  
 (10b) (ITA): alla sinistra della candela **ho** un tubetto di dentifricio  
 senza tappo  
 (10c) (GER): daneben **hab** ich ne küchenpapierrolle

Verbless construction:

- (11a) (FRE): et un petit peu en dessous un pot bleu  
 (11b) (ITA): e accanto una mela  
 (11c) (GER): daneben ein apfel

As shown in Table 6, German differs from the two Romance languages in the huge amount (53%) of verbless constructions. As far as one can tell, the NPs referring to the entities are in the nominative (due to syncretism this can only be seen with masculine NPs as in [11c], but there is no exception with these). When these structures are analysed as elliptical, they should thus be considered as instances of *be* constructions (with the Entity as the subject NP) rather than *have* constructions, where the Entity would have to show up as an (accusative) object NP.

Disregarding the verbless utterances for a moment, we will next study the distribution of the verb-containing constructions across the three languages. Figure 1 shows the relative frequency of the main construction types in verb-containing utterances.

The two Romance languages show a comparable distribution of construction types amongst the verb-containing utterances: next to the *there\_be/have* construction, the *have* construction is used in the majority of the cases. German differs from the two Romance languages not only in the use of the *be* construction instead of *there\_be/have* construction, but also in the comparably low frequency of the *have* construction. The predominance of the *be* construction in German corresponds to what was found in picture description monologues (Carroll & von Stutterheim, 1997). Recall that the verbless cases (Table 6) are likely elliptical *be* constructions as well.

Some of these results were expected on the basis of the typological differences described in Table 2. The main function of Introductions is a localisation going from the Relatum to the newly introduced and focal Entity-Theme. Speakers chose Inverted Locational Predications (ILP) to realize this function. Purely existential *es gibt* is thus impossible in German, and German speakers resort to the *be* construction instead. *There-be/have* constructions in French and Italian cover ILP-localisations and can therefore be used, while the *be* construction is not available for this function (*\*a gauche est une orange* 'to the left is an orange').

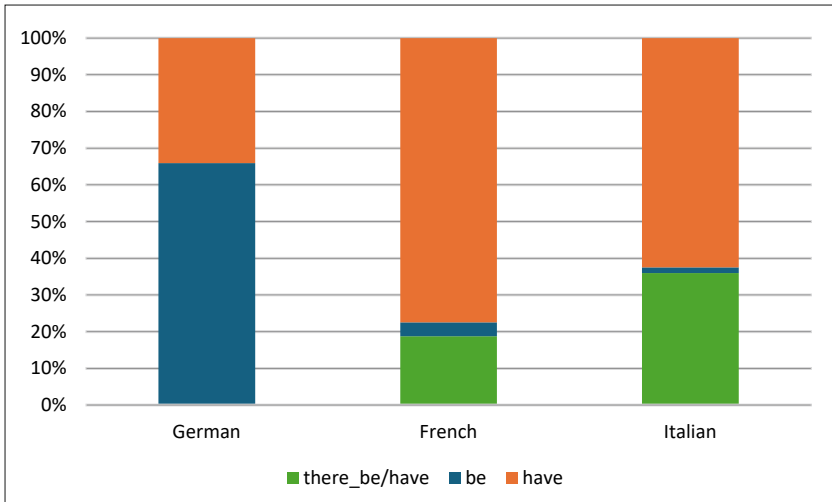


Figure 1. Frequency of construction types (First Move Introduction; only verb-containing utterances)

This typological distinction does, however, not explain the unexpected differences concerning the frequency of *have* constructions in German and in the Romance languages. Recall that *have* constructions were not attested in monological picture description tasks involving only one picture (e.g., Watorek, 2003). We assume that the structure shows up in our dialogue data as it presents an easy way to encode the speaker's Frame of Reference. Moreover, like the ILP constructions *there\_have/be* (FRE, ITA) and *be* (GER), it allows an introduction that goes from the Relatum in initial position to the Entity-Theme in focal position: *links habe ich eine orange* (GER); *à droite j'ai une orange* (FRE); *a destra ho un'arancia* (ITA). We will come back to this issue in the following section where we consider the consequences of these cross-linguistic differences for the integration of the optional information units Location and Frame of Reference as well as the information structure of the resulting utterances.

### Information Units

In a first step, we only count whether the relevant information units are present or not. In addition to the obligatory Entity (ENT), i.e., all types of NPs and pronouns referring to the entities on the pictures, we will consider localising adverbials or prepositional phrases as Location (LOC)<sup>13</sup> and all kinds of

13 The data contained also temporal adverbials (FRE *ensuite* / ITA *poi* / FRE *dann*) that can be interpreted as conveying spatial information, when speakers follow a linear

expressions referring to the current speaker's picture or the speaker as its possessor as Frame of Reference (FRA).

**Location** Recall that the items on the stimulus pictures were presented in a random arrangement. Speakers could thus not rely on world knowledge when localising entities, because no kind of spatial relation between objects could be a priori excluded. When speakers followed a global strategy, regions of the picture were mainly indicated with adverbials like *bottom left* without mentioning the global Relatum (i.e., the picture). In a point-by-point strategy, expressions referring to specific locations typically had the form of a prepositional phrase including reference to the Relatum (*under the teapot*) but there were some (language specific) variants as well. German has so-called prepositional adverbs (*darunter* – 'there-below', *daneben* – 'there-to-the-side') in which the first part (*da*) anaphorically refers to the space occupied by the Relatum, and the second part indicates the relative position of the space occupied by the Theme. In French and Italian, reference to an already established entity functioning as a Relatum can be elliptical (ITA: *in alto a sinistra c'è un'arancia... sotto* [all'arancia] *c'è un bicchiere di plastica... di fianco* [al bicchiere] *c'è una pera*. 'On top to the left there is an orange below [the orange] there is a plastic cup... besides [the cup] there is a pear.')

**Frame of Reference** Mentioning this information unit in a First Move Introduction might seem redundant, as each speaker could only see his / her own picture, but quite a few introductions indeed contain this information. Expressions indicating the Frame of Reference are quite heterogeneous. We subsume all verbal cues indicating that a speaker is referring to one of two pictures under this category. The most prominent solution consisted of an explicit reference to the speaker's picture (FRE: *dans mon image*; ITA: *nella mia immagine*) or to the location of the speaker (GER *bei mir*)<sup>14</sup>. Moreover, reference to the speaker is always present in the syntactic frame of *have* constructions that opens a second argument slot for a 1st person pronoun. Because of the morphological differences described in Table 1, German, French and Italian behave differently in this respect. As a pro-drop language,

---

strategy and engage in a mental tour through the picture, as indicated, for example, by an explicit indication of the direction taken (*I'm going from left to right*). In these cases, it is likely that the order of mention corresponds to the relevant spatial relations. As it was not always possible to decide whether these temporal adverbials were indeed encoding spatial information or the temporal organisation of discourse segments, they were not counted as expressions of Location.

- 14 Speakers can also produce possessive pronouns to indicate that they are talking about one of two similar items (*my pear is below the orange*). As this definite marking presupposes an agreement on the existence of the relevant entity, it occurred only rarely in introductions and was therefore not counted as an expression of Frame of Reference.

a subject pronoun (*io*) is not obligatory in Italian; however, even with a null subject, reference to 1st person is available through verb morphology (*ho*: have.1P.Sg.). As non-prodrop languages, German and French always have a subject pronoun in *have* sentences; this is a clitic in fixed preverbal position in French (*je*), while German *ich* can be stressed and is syntactically mobile. French has the further possibility of a strong pronoun (*moi*) occurring as an independent constituent, together with the clitic pronoun.

Given that *have* constructions were not attested in description tasks involving only one picture, their presence can be interpreted as reflecting the speaker's decision to encode the Frame of Reference in a relatively light way. This strategy can be reinforced by the other markers (overt pronoun in ITA, strong pronoun in FRE, adverbials referring to the picture in all Ls), whereby the latter can also be used outside of the *have* construction. Table 7 shows the relative frequencies of mention for the information units Localisation and Frame of Reference in First Move Introductions. Cells with an absolute number of utterances below 3 are shaded. As reference to the speaker is inherent to the *have* constructions, the relevant cells (marked with\*) show a 100% in all languages: see below for further discussion.

Table 7. Absolute numbers and frequency of Localisation and Frame of Reference (First Move Introduction)<sup>15</sup>

		<i>there_be/ have</i>	<i>be</i>	<i>have</i>	Tot verb- containing	<i>verb- less</i>	Total
German	LOC	-	29 (100%)	15 (100%)	44 (100%)	21 (42%)	65 (69%)
	FRA	-	14 (48%)	15 (100%)*	29 (13%)	0	29 (31%)
	<i>Tot</i>	0	29	15	44	50	94
French	LOC	15 (100%)	2 (67%)	55 (89%)	72 (90%)	2 (50%)	74 (88%)
	FRA	1 (7%)	0	62 (100%)*	63 (79%)	0	63 (75%)
	<i>Tot</i>	15	3	62	80	4	84

15 As in all the following tables, percentages are rounded. The percentages in grey correspond to 3 or less utterances.



Table 7. (continued)

		<i>there_be/ have</i>	<i>be</i>	<i>have</i>	Tot verb- containing	verb- less	Total
Italian	LOC	14 (61%)	1 (100%)	24 (60%)	39 (61%)	6 (40%)	45 (57%)
	FRA	3 (13%)	0	40 (100%)*	43 (67%)	0	43 (54%)
	<i>Tot</i>	23	1	40	64	15	79

**Location** can be explicitly expressed in all construction types, but it occurs with reduced frequency in verbless constructions. When only verbal constructions are considered, both German and French speakers nearly always mention LOC, while Italian speakers are less systematic (90–100% of mention in French / German vs. 61% in Italian). Introductions without LOC also appear in the *there\_be* construction in Italian. These cases are mainly due to a particular strategy adopted in some dialogues. When the interlocutors do not immediately realise that not only the type of entities, but also their location can differ between the pictures, they sometimes start with a list of Entities produced by one speaker. Localisations are only provided in a later segment of the discourse, typically after the first mismatch that informs speakers about the possibility of differences in LOC. A single pair of interlocutors starting with the assumption that only the Entities matter, can severely alter the quantity of LOC in First Move Introductions. The lower frequency of LOC expressions in Italian does probably not reflect structural properties but is simply due to more Italian dialogue pairs falling into this trap.

**Frame of Reference:** The first observation concerns the absence of FRA markings in verbless introductions across all languages. The comparably high totals for FRA in French and (partly) Italian are directly reflecting the low number of verbless introductions in these languages. Put the other way around, one could also interpret the high number of verbless utterances in German as a way to compensate for the missing *there\_be/have* structure. This structure allows speakers of French and Italian to introduce entities without additional specification of LOC or FRA (*c'è una mela*), whereas the *be* construction that is predominant in German requires at least one of these information units. In that sense, the verbless constructions might be interpreted as a variant of the *be* constructions that allows to circumvent this constraint.

In the Romance languages, the (very low frequent) *be* construction seems incompatible with FRA markings, whereas this is not the case in German,

where 45 % of the introductions in *be* constructions contain an adverbial FRA marker (exclusively *bei mir*).

Reference to the speaker is inherent to the *have* constructions: therefore, the relevant cells show a 100 % of FRA in Table 7. This masks, however, the additional presence of optional (and stronger) FRA-markings. In Table 8, we indicate how often the *have* constructions in the three languages contained one of the following additional and optional FRA-markers: Overt 1st person pronoun *io* in Italian; strong pronoun *moi* in French, adverbials referring to picture or speaker in all languages.

Table 8. Absolute numbers and frequency of optional FRA-Markers in the *have* construction (First Move Introduction)

	<i>Have_construction with optional FRA-marking</i>	<i>Total</i>
<b>German</b>	0	15
<b>French</b>	16 (26%)	62
<b>Italian</b>	18 (45%)	40

Even though FRA-adverbials are not excluded from *have* constructions (*In my picture I have a tea pot*), they are nearly absent in the three languages (there is only a single occurrence in French, and it is combined with the strong pronoun). Nearly all additional FRA markers in Table 8 are thus pronouns. Speakers of Italian used overt pronouns in 45% of the cases. This relatively light marking seems to be licensed under First Move conditions. The strong pronoun (*moi*) in French that occurs in combination with the obligatory clitic pronoun leads to a construction that is heavier than the corresponding one in Italian. This might explain its lower frequency (26%). The German subject pronouns (that were not counted as additional FRA markers because of their obligatory presence) were unstressed and the vast majority occurred in postverbal position.

To sum up, all verb-containing introductions in German contain a LOC expression. The many verbless constructions drop this rate to 66 % overall. LOC expressions are equally frequent in verb-containing utterances in French. Due to the lower number of verbless constructions, their overall share is even higher than in German. The lower frequency of LOC in Italian introductions cannot be explained by structural differences and is attributed to the comparably high number of Italian interlocutors starting with a list of entities before realising that differences between the pictures can also be a matter of location.

FRA is marked in only 30% of the German introductions, mainly by adding a *bei mir* adverbial to the predominant *be* construction, and more rarely by choosing the *have* construction instead. FRA is clearly more frequently marked in French and Italian. There are, however, only 7–13% FRA markings in the *there\_be/have* constructions. The bulk of the markings is due to the predominant *have* constructions in the Romance languages. They have an inherent FRA component and are often used with additional pronouns whose frequency might depend on their perceived heaviness.

### Information Structure

With respect to the position of the different information units, the Entity occurs in final, i.e., focal position, in the large majority of cases in all three languages. This holds irrespectively of the chosen construction and the number of information units involved. In this section, we are particularly interested in the filling of the initial utterance positions that link the introductions to the preceding context and represent the speakers' choice for a topical backbone organising their discourse. As we will see later, this is particularly relevant in comparison to the Second Move Introductions.

In the following, we will report which information units (FRA or LOC) were placed in a position preceding the finite verb. We are therefore only considering verb-containing utterances. Due to the V2-constraint in German, this position is in principle restricted to only one constituent (either FRA or LOC), whereas a combination is possible in French and Italian (both SVO languages). Since we were interested in speakers' choices, elements that obligatorily occur in preverbal position were excluded from the analyses (we will thus count the personal pronouns *io*, *ich*, and *moi*, but not the French clitic *je* that has to directly precede the verb in all cases). Furthermore, each information unit was counted only once, even though double markings (*moi* [= FRA], *dans mon image* [= FRA], *j'ai une poire*) occasionally occurred. Utterances in which the preverbal position was not used for reference to either LOC or FRA (e.g., the preverbal position was not filled at all, or it was filled with an NP referring to the Entity, or with a temporal adverbial, for example), were counted as "other". The results are summarised in Table 9. The high number of verbless utterances in German and (less so) Italian, that were excluded from this analysis, explains the differences in the totals.

Table 9. Absolute numbers and frequency of Information units in preverbal position (First Move Introduction)

		<i>there_be/ have</i>	<i>be</i>	<i>Have</i>	<b>Total</b>
<b>German</b>	<b>LOC</b>	–	22 (76%)	11 (73%)	33 (75%)
	<b>FRA</b>	–	3 (10%)	2 (13%)	5 (11%)
	<b>both</b>	–	0	0	0
	<b>other</b>	–	4 (14%)	2 (13%)	6 (14%)
	<i>Tot</i>	0	29	15	44
<b>French</b>	<b>LOC</b>	14 (93%)	2 (67%)	34 (55%)	50 (62,5%)
	<b>FRA</b>	0	0	2 (3%)	2 (2,5%)
	<b>both</b>	1 (7%)	0	14 (23%)	15 (19%)
	<b>other</b>	0	1 (33%)	12 (19%)	13 (16%)
	<i>Tot</i>	15	3	62	80
<b>Italian</b>	<b>LOC</b>	13 (57%)	1 (100%)	15 (37,5%)	29 (45%)
	<b>FRA</b>	2 (9%)	0	8 (20%)	10 (16%)
	<b>both</b>	1 (4%)	0	8 (20%)	9 (14%)
	<b>other</b>	7 (30%)	0	9 (22,5%)	16 (25%)
	<i>Tot</i>	23	1	40	64

Figure 2 displays the relative frequency of the relevant information units in preverbal and topic position in the three languages.

LOC is topical in most cases (in French and Italian, either alone or together with FRA), although in different proportions: 75% GER; 81% FRE; 59% ITA. In comparison, FRA is much more rarely topical (once again, in French and Italian it can occur together with LOC): it is chosen as a topic in 11% GER, 21% FRE, 30% ITA. In other words, as a single topic, LOC is preferred over FRA. The possibility to combine LOC and FRA encourages speakers of French and Italian to also mention FRA early in their utterances, whereas the need to decide between LOC and FRA in topic position in German clearly privileges LOC over FRA.

Concerning the verbless structures that we have excluded here, the linear order of information units is also telling. In German, for example, where

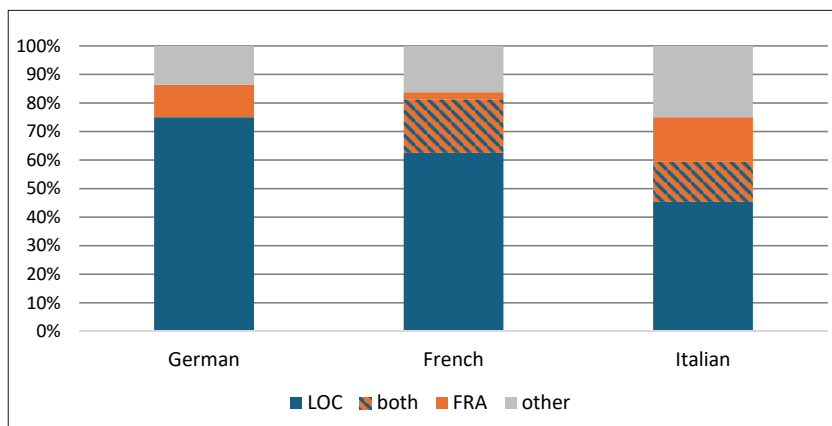


Figure 2. Frequency of Information units in preverbal position (First Move Introduction; only verb-containing utterances)

verbless utterances make up for more than half of the First Move Introductions, they typically have either LOC or some other constituent like a temporal adverbial preceding the NP referring to the Entity (if there is anything at all). Both elements can also be combined, as in: *und dann rechts etwas höher so 'ne leere Klopapierrolle* ('and then to the right, somewhat higher, an empty toilet roll'). However, such combinations account for only a small fraction of the verbless utterances and speakers of German never combined FRA and LOC in the positions preceding ENT. It is therefore unlikely that they resort to verbless utterances in order to circumvent the V2-constraint and find a structure resembling the Romance languages with their possibility of topicalising both information units. Instead, the verbless constructions in German can be considered a reduced variant of the frequent and multifunctional *be* construction, albeit without the obligatory extra information unit (FRA or LOC or both) associated with the latter.

## Results Part II: Second Move introductions

As explained in the introduction of the paper, we will now analyse the second turns of interactional sequences, i.e., utterances that directly react to what the interlocutor said before. Second Move Introductions were found in replies to questions and in contexts, in which speakers introduce an entity that differs from the entity their interlocutor proposed for a given location (A: *Below I have a teapot.* B: *No, there is an apple in my picture!*). The latter type will be referred to as Contrastive Introductions. They will constitute the focus of

our analysis after a short paragraph on the replies. Table 10 summarises the frequency of both types in the three languages.

Table 10. Second Move Introductions

	Total Second Move Introductions	Introductions in replies	Contrastive Introductions
German	28	10	<b>18</b>
French	25	10	<b>15</b>
Italian	27	12	<b>15</b>

### Introductions in replies to questions

Replies can again come in two forms. The majority of them are replies to explicit WH-questions (*What do you have below the teapot?*). The replies are often elliptical, i.e., they are verbless and consist of the NP referring to ENT only. There are also some verb-containing replies that typically have LOC in preverbal topic position. The other type of replies are so-called *co-constructions* (Sacks, 1995; Lerner, 1991). These are cases in which both speakers co-construct an utterance consisting of the topical location provided by speaker A, followed by the verb (if there is one) and the Entity produced by Speaker B, as in example (12). Replies to WH-questions outweigh co-constructions in all languages.

(12) A: *Und dadrunter?* B: *ist eine Klopapierrolle.*

Introductions in replies will not be further considered in the analyses of Second Move Introductions.

### Contrastive Introductions

In the following analyses, we include all utterances specifying the presence of a contrasting entity in a position that was mentioned in the preceding First Move Introduction. The relevant entities were often new, i.e., mentioned for the very first time in this contrastive utterance (A: *next to the orange there is a teapot.* B: *I have a candle there!*). In some cases, however, the relevant entity was mentioned by the current speaker or the interlocutor at some earlier point in the dialogue and is taken up again when specifying that there is a contrast between both pictures concerning the entities encountered in a similar location. In accordance with all other introductions, the focus is on the expression referring to the entity in both cases. As with the First Move

Introductions, we will treat the construction types chosen in Second Move Introductions, the integration of the information units LOC and FRA, and the information structure of the utterances.

### Construction types

As can be seen from Table 11, the number of central construction types used is reduced to two (the *be* construction in German and the *have* construction in French and Italian). The *there\_be/have* construction and the verbless construction play a marginal role at best.

Table 11. Absolute numbers and frequency of construction types (Second Move Introduction)

	<i>there_be/have</i>	<i>Be</i>	<i>have</i>	verbless	<i>Total</i>
<b>German</b>		13 (72%)	3 (17%)	2 (11%)	<b>18</b>
<b>French</b>	0	1 (7%)	14 (93%)	0	<b>15</b>
<b>Italian</b>	0	0	14 (93%)	1 (7%)	<b>15</b>

Examples for the two main construction types:

*be* construction:

- (13) (GER): A: *rechts daneben ist bei mir ne birne*  
 B: *da ist bei mir ein tennisball.*

*have* construction:

- (14a) (FRE): A: *juste à gauche j'ai une balle de tennis*  
 B: *non moi juste à gauche j'ai un (.) un tube de dentifrice.*  
 (14b) (ITA): A: *di fianco c'è una pera*  
 B: *no (.) io di fianco al bicchiere ho una pallina da tennis.*

The considerable reduction of verbless constructions in Second Move Introductions does not come as a surprise. As speakers are engaged in contrastive comparisons, mentioning the Frame of Reference is important (see details below) and verbless constructions are not suitable to host this information unit as we saw above. The motivation for abandoning the *there\_be/have* construction is less clear, as it can in principle be combined with FRA-expressions in French and Italian as shown in Table 7. We will come back to this point in

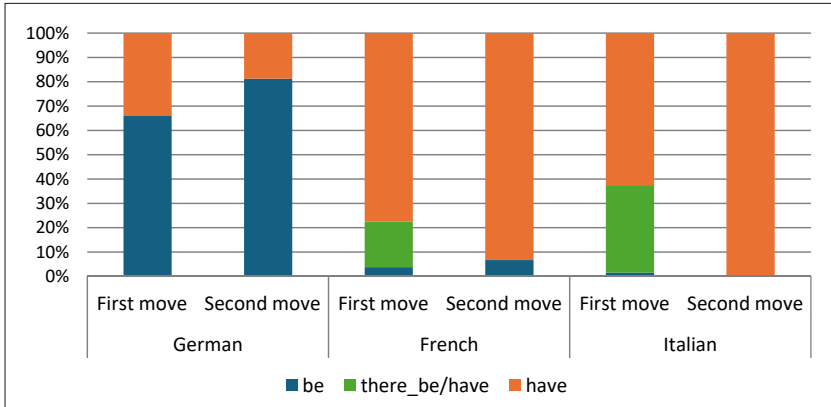


Figure 3. Distribution of construction types in First vs. Second Move Introduction (only verb-containing utterances)

the discussion and first home in on the relative distribution of verb-containing constructions. Figure 3 shows, whether and how the speakers’ choices differ from their preferences in First Move Introductions.

In Second Move Introductions, speakers of all three languages select the construction type that was already dominant in First Move Introductions, now largely neglecting all other possibilities. Within the remaining two syntactic constructions, German and Romance languages are once again different: At the expense of variation, German uses *be* as the preferred pattern, while the Romance languages systematically chose *have*. In the next steps, we will check whether this is likely to be motivated by differences in the presence of the information units LOC and FRA and by the information structure of the relevant utterances.

Information Units

Table 12 shows the relative frequencies of mention for the information units Localisation and Frame of Reference in Second Move Introductions.<sup>16</sup>

Figure 4 presents a comparison of the percentages across all syntactic constructions in First Move Introductions (Table 7) and Second Move Introductions (Table 12). It reveals that the relevance of mentioning FRA has increased in all languages, whereas the frequency of mentioning LOC has decreased in French and Italian but increased in German.

The result for FRA is expected. Although every speaker can still only see and comment on their own picture, FRA constitutes contrastive information in Second Move Introductions. In German, the increase is due to the *bei mir* adverbial that now occurs in 100% of the highly frequent *be* constructions.

16 The percentages in grey correspond to 3 or less utterances.



Table 12. Absolute numbers and frequency of Localisation and Frame of Reference (Second Move Introduction)

		<i>there_</i> <i>be/</i> <i>have</i>	<i>be</i>	<i>have</i>	Tot verb- containing	verb- less	Total
German	LOC	–	13 (100%)	2 (100%)	15 (100%)	0	15 (83%)
	FRA	–	13 (100%)	2 (100%)	15 (100%)	0	15 (83%)
	<i>Tot</i>	0	13	2	15	3	18
French	LOC	–	0	8 (57%)	8 (53%)	–	8 (53%)
	FRA	–	1 (100%)	14 (100%)	15 (100%)	–	15 (100%)
	<i>Tot</i>	0	1	14	15	0	15
Italian	LOC	–	–	5 (36%)	5 (36%)	0	5 (33%)
	FRA	–	–	14 (100%)	14 (100%)	0	14 (93%)
	<i>Tot</i>	0	0	14	14	1	15

In the Romance languages, the increase is a consequence of the increase of *have* constructions that come with an implicit FRA. The contrastiveness of FRA becomes clearer when we have a closer look at these constructions and account for additional and optional FRA-markings (overt/strong pronouns, adverbials), as we did for First Move Instructions in Table 8. Table 13 compares First and Second Move Introductions.

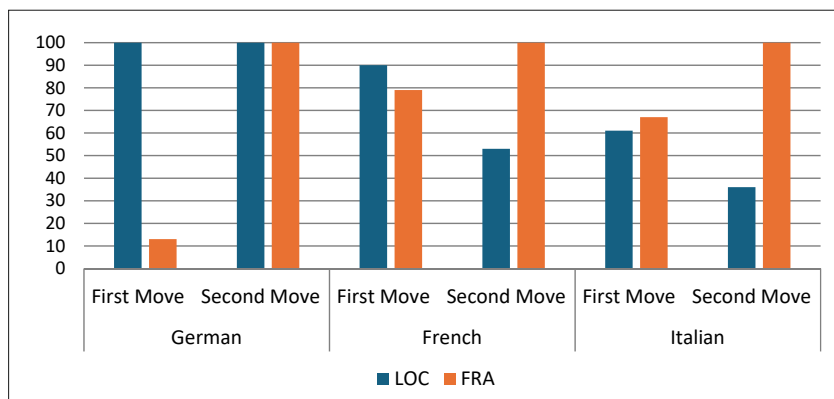


Figure 4. Frequency of Localisation and Frame of Reference in First vs. Second Move Introduction (only verb-containing utterances)

Table 13. Absolute numbers and frequencies of additional FRA-Markers in the *have*\_construction in First vs. Second Move Introduction

	<i>Have_construction with optional FRA-marking</i>			
	<b>First Move Introduction</b>	<i>Total</i>	<b>Second Move Introduction</b>	<i>Total</i>
<b>German</b>	0	15	0	2
<b>French</b>	16 (26 %)	62	12 (86 %)	14
<b>Italian</b>	18 (45 %)	40	13 (93 %)	14

As before, mainly pronouns were used as additional FRA-markers. Due to the missing variability in the German pronoun inventory, the *have* construction that was already not very frequent in First Move Introductions is even less preferred for expressing the contrast in the FRA unit that is needed for Second Move Introductions. In the Romance languages, on the other hand, there is a clear increase in the frequency of optional additional FRA markings in Second Move Introductions. Speakers thus increased the use of the *have* construction and within that the use of additional FRA markers to solve the communicative need of marking a contrast in this information unit.

The results for the decrease of LOC (see Figure 4) that was attested in the Romance languages is equally expected. In the contrastive introductions, speakers mention different entities that are situated at the same position (on both pictures) and reference to the Relatum can thus often be maintained from an earlier utterance. The high frequency of LOC in German rather comes as a surprise. We will return to this finding in the following section.

### Information Structure

In Second Move Introductions, the expression encoding the Entity is again focal and thus mainly placed at the end of the relevant utterances. The Location is given, and the Frame of Reference is contrastive. In the following we will report which information units (FRA or LOC) were placed in a position preceding the finite verb. We are therefore again considering verb-containing utterances only and we excluded elements that obligatorily occur in preverbal position (the French clitic *je*). See the description preceding Table 9 for further coding details. The results are summarised in Table 14. Examples for utterances with LOC and/or FRA in preverbal position are given below.

Table 14. Absolute numbers and frequency of information units in preverbal position (Second Move Introduction)

		<i>there_be/have</i>	<i>Be</i>	<i>have</i>	Total
German	LOC	–	10 (77%)	1 (50%)	11 (73%)
	FRA	–	2 (15%)	1 (50%)	3 (20%)
	both	–	1 (8%)	0	1 (7%)
	other	–	0	0	0
	<i>Tot</i>	0	13	2	15
French	LOC	–	0	2 (14%)	2 (13%)
	FRA	–	1 (100%)	6 (43%)	7 (47%)
	both	–	0	6 (43%)	6 (40%)
	other	–	0	0	0
	<i>Tot</i>	0	1	14	15
Italian	LOC	–	–	0	0
	FRA	–	–	8 (57%)	8 (57%)
	both	–	–	5 (36%)	5 (36%)
	other	–	–	1 (7%)	1 (7%)
	<i>Tot</i>	0	0	14	14

(15a) GER: **da ist bei mir** eine birne

(15b) GER: **rechts neben dem becher bei mir** ist die birne

(16) FRA: **de façon horizontale sur la gauche moi** j'ai une euh euh orange

(17) ITA: **io invece nella seconda partendo da destra** ho la pallina da tennis

Note that the one occurrence of a German *be\_* construction with both LOC and FRA in preverbal position (see example 15b) would be ungrammatical in a written register whereas utterances with a complex topic component regularly occur in French and Italian (examples 16 and 17). Figure 5 presents the findings for the preverbal position in First and Second Move Introductions in a comparative perspective.

The information unit FRA changes from *shifted* in First Move Introductions to *contrastive* in Second Move Introductions. As a consequence, the frequency

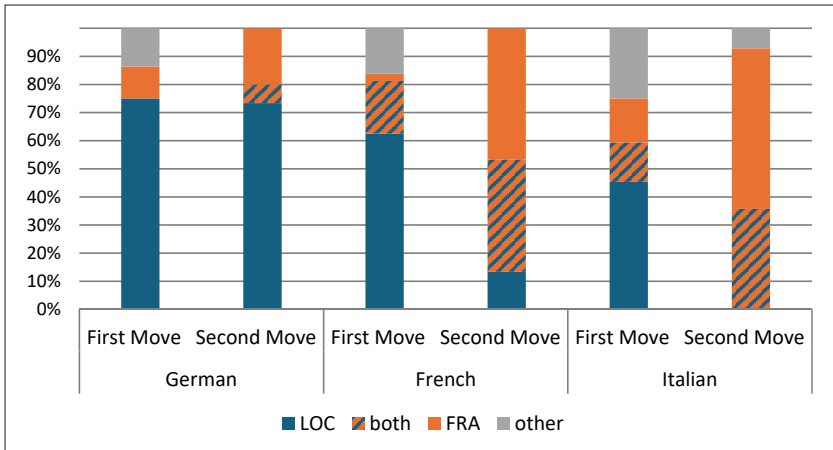


Figure 5. Information Units in preverbal position in First vs. Second Move Introduction

of mention of FRA in preverbal position sharply increases in French and Italian, in particular when double topics (“both”) are considered. In German, on the other hand, there is only a small difference between First and Second Moves. The topic position is mostly occupied by a LOC expression and the changing information status of this information unit is barely relevant for word order. This could be a reflex of the V2 constraint that prevents German speakers from placing both information units in preverbal position. However, a closer look at the preferences in the Romance languages speaks against this idea. When there is only one preverbal constituent in the Romance languages, LOC alone is rare in French and not used at all in Italian Second Moves, whereas FRA alone is frequent in both languages.

German speakers prefer LOC as a topic, and they stick to this decision in First Move and Second Move Introductions. In the Second Move Introductions, choosing between LOC and FRA corresponds to choosing between a Continuation Topic and a Contrastive Topic. There seems to be a clear preference for the former in German, and for the latter (or a combination of both) in the Romance languages. This suggests that a textual preference, rather than only a syntactic constraint, is at play: German speakers prefer building discourse cohesion on continuity (anaphoric or given LOC), while speakers of French and Italian prefer to encode the contrasting information unit FRA as (part of) the topic in Second Moves.

## Summary and Conclusions

The aim of our paper was to explore how native speakers of German, Italian and French introduce entities in an oral picture description task involving a dialogic dimension and a comparative goal (identify an unknown number of differences between two pictures). This makes it relevant for speakers to refer to their own picture (FRA) and to introduce entities in a contrastive way when they react to the interlocutor's earlier statements. The results show both (a) the influence of the type of task and (b) cross-linguistic differences concerning the preferred means employed for strengthening text cohesion under the given conditions.

In First Move Introductions, speakers of Romance languages clearly prefer to introduce entities with *have* constructions instead of the existential *there\_be/have* constructions (*il y a / c'è*). *Have* constructions are also present in the German data, albeit to a lesser extent. German speakers frequently use *be* constructions (or equivalent verbless constructions) instead. The use of possessive *have* in the three languages is in sharp contrast with the findings of previous research on monologic descriptive discourse (Carroll et al., 2000; Watorek, 2003; von Stutterheim & Carroll, 2018). We assume that its presence is motivated by the task implying the existence of differences between the two images. We conclude that *have* constructions are licenced as a relatively “light” option, allowing speakers of all three languages to signal a shift between two Frames of Reference without, however, expressing contrasts between the pictures as they do not know yet whether the interlocutor's picture contains the entity they are introducing. Across languages and construction types, there is a general tendency to fill the topical position in First Move Introductions with the information unit LOC. The *Quaestio* “What is there in L?” that Watorek (2003, p. 330) proposed for static spatial descriptions in French seems thus to be valid for Italian and German as well.

The differences between Romance languages, on the one hand, and German, on the other, becomes more clear-cut in Second Move Introductions, where the information unit Frame of Reference (FRA) is contrastive instead of being only shifted. Speakers of French and Italian altogether abandon the *there\_be/have* constructions, even though these can in principle be combined with FRA-expressions. The *have* construction that is already frequent in the First Move Introductions turns out to be particularly suitable for encoding this additional information unit, as speakers can easily add optional overt (Italian) or contrastive pronouns (French) to express the contrast in an utterance initial position. The presence of optional pronouns as overt markers of FRA contrasts does not hinder the expression of location (LOC) in the preverbal position. Recall that Location is maintained and non-contrastive information in Second Moves because speakers talk about different entities located in a position

that can be identified as being *the same location on different pictures*. This non-contrastive Localisation can be achieved with a global description like *the upper left corner* or with a local description like *next to the orange*, when both interlocutors have agreed upon the position of the Relatum (*orange*) in the preceding discourse. In the Romance languages, the left periphery of an utterance can thus be used to signal contrast and discourse continuity at the same time (*moi juste à gauche j'ai un tube de dentifrice*).

Optional pronouns do not have a counterpart in German and stressed variants of the personal pronoun *ich* were not attested in the context of entity introductions. This might be one of the reasons why German speakers only rarely use *have* constructions in First Move Introductions, and almost completely avoid them in Second Move Introductions when a contrast on the Frame of Reference is at play. Combinations of *have* with FRA adverbials are in principle possible but barely ever occurred. The personal pronoun *ich* is more prominent than the French clitic or the Italian null pronoun, so speakers presumably deemed this double marking (*bei mir habe ich ...*) too heavy and chose the *be* construction instead.

Furthermore, the V2 constraint forces German speakers to choose between LOC and FRA in the topical preverbal position. Our results show that they invariably select LOC, i.e., a topic of continuity, rather than contrastive FRA for this position (*da ist bei mir Zahnpasta*). The optional *bei mir* adverbials increase in the Second Moves, but FRA does not become the starting point of the utterance, even though syntactic constraints would not hinder speakers to start with *bei mir* and put the local adverbials in a postverbal position instead. Given the necessity of a choice, the initial and topical part of the sentence is used for continuity purposes and the local information constitutes the backbone of coherence in both First and Second Move introductions in German. This also holds for the comparably high number of verbless utterances.

Overall, we found that the adaptation to an interlocutor in a jointly constructed picture description dialogue has different types of consequences, compared to information management in a monologic situation. The three factors discussed in the Introduction left language specific traces in our data. The necessity for a **collaborative discourse construction** leads to local sequences of First and Second Moves. In Second Move Introductions, speakers have to indicate whether their contribution is a confirmation or a disconfirmation of their interlocutor's statement. In the latter case (which was in the centre of our interest) they also have to make clear which part of the information is congruent with the interlocutor's description and where the information conveyed is in contrast to what was said before. In doing so, speakers have to cope with an **additional information unit** (the Frame of Reference). Next to Inverted Locational Predications (ILPs) that were expected as suitable constructions for the introduction of Entities, speakers of all languages solved

this problem with the help of possessive constructions where reference to the speaker is built into the argument structure of *have*. The rich pronoun inventory of the Romance languages reinforces this preference, whereas the multifunctional *be* construction that can be combined with FRA adverbials is prominent in German. The importance of the pronoun inventory in the Romance languages vs. the use of adverbials, particles and intonation in German ties in with earlier findings on signalling discourse alternatives in narratives (Benazzo & Andorno, 2010; Dimroth et al., 2010; Benazzo et al., 2012).

The need to signal **contrastive content** had an impact on information structure. In Second Move Introductions, the information units Entity and Frame of Reference are in contrast to the preceding First Moves, whereas the Location is maintained. At the same time, only ENT is clearly focal, whereas LOC and FRA are indicating what is talked about, i.e., candidates for the topic component of the utterances. In the case of ENT (new and focal) as well as LOC (maintained and topical) these properties go well together, and word order reflected the information status of these information units in all three languages (preverbal LOC, postverbal ENT). The information unit FRA (contrastive and topical) was therefore particularly interesting, and speakers solved this problem in language specific ways. In the Romance languages, FRA was the dominant topic in Second Move Introductions, sometimes in combination with LOC (*Moi (à droite) j'ai une orange*). In German, the topic position was nearly exclusively filled with LOC (*Da ist bei mir eine Orange*). A contrastive information unit was thus deemed to be a suitable way of cohesion building in French and Italian Second Move Introductions, whereas German speakers clearly preferred to use the topic position for signalling continuation.

In languages with SVO syntax, the *have* construction comes with an obligatory preverbal slot for an empty or clitic pronoun that can easily be made more prominent by open or strong pronoun forms in the same position. In German (V2), *be* constructions can be used in First and Second Move Introductions. In the absence of a subject first constraint, the preverbal position is generally filled with the LOC unit that builds the cohesive backbone of the dialogues. Contrastive FRA, if needed in Second Move Introductions, is realised as an adverbial in postverbal position instead.

Coming back to the variant of the *Quaestio* underlying static spatial descriptions in German, French, and Italian (*What is there in L?*), the specific conditions of our dialogue task induce the presence of an alternative *Quaestio*. First Move Introductions are not only descriptions, they also transport an inherent invitation to the interlocutor to check whether the relevant spatial configuration is similar on his or her picture (“What about your picture?”). Differences in information structure in Second Move Introductions between the two language groups suggest that speakers answer two slightly different variants of this question. While speakers of German still formulate their Second Move

Introductions as an answer to “What do you see in L instead?”, with the maintained LOC as a topical starting point, speakers of French and Italian adopt a more explicitly contrastive starting point “What do you see in your picture instead?”. The specificities observed by Christiane von Stutterheim concerning the dominance of a spatial perspective for German arise in dialogue descriptions as well.

## References

- Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., et al. (1991). The HCRC map task data. *Language and Speech*, 34(4), 351–366.
- Benazzo, S., & Andorno, C. (2010). Discourse cohesion and Topic discontinuity in native and learner production. Changing topic entities on maintained predicates. *EUROSLA Yearbook 10* (pp. 92–118). Benjamins.
- Benazzo, S., Andorno, C., Patin, C., & Interlandi, G. (2012). Perspective discursive et influence translinguistique: Exprimer le contraste d’entité en français et en italien L2. *Language, Interaction, and Acquisition*, 3(2), 173–201.
- Carroll, M., & von Stutterheim, C. (1993). The representation of spatial configurations in English and German and the grammatical structure of locative and anaphoric expressions. *Linguistics*, 31(6), 1011–1041.
- Carroll, M., & von Stutterheim, C. (1997). Relations entre grammaticalisation et conceptualisation et implications sur l’acquisition d’une langue étrangère. *Acquisition et Interaction en Langue Étrangère (AILE)*, 9, 83–116.
- Carroll, M., Murcia-Serra, J., Watorek, M., & Bendiscioli, A. (2000). The Relevance of Information Organization to Second Language Acquisition Studies: The Descriptive Discourse of Advanced Adult Learners of German. *Studies in Second Language Acquisition*, 22(3), 441–466.
- Creissels, D. (2019). Inverse-locational predication in typological perspective. *Italian Journal of Linguistics*, 31(2), 37–106.
- Cztinglar, C. (2002). Decomposing existence: Evidence from Germanic. In A. Werner, & J.-W. Zwart (Eds.), *Issues in formal German(ic) typology* (pp. 85–126). Benjamins.
- Dimroth, C., Andorno, C., Benazzo, S., & Verhagen, J. (2010). Given claims about new topics. How Romance and Germanic speakers link changed and maintained information in narrative discourse. *Journal of Pragmatics*, 42(12), 3328–3344.
- Dimroth, C., & Starren, M. (2022). The impact of grammar on the construal of discourse alternatives in German and English. *Languages* 2022, 7, 240. doi.org/10.3390/languages7030240
- Klein, W. (2008). The topic situation. In B. Ahrenholz et al. (Eds.), *Empirische Forschung und Theoriebildung: Beiträge aus Soziolinguistik, Gesprochene-Sprache- und*



- Zweitspracherwerbsforschung: Festschrift für Norbert Dittmar* (pp. 287–305). Lang.
- Klein, W., & von Stutterheim, C. (1987). Quaestio und referentielle Bewegung in Erzählungen. *Linguistische Berichte*, 109, 163–183.
- Koch, P. (2012). Location, existence, and possession: A constructional-typological exploration. *Linguistics*, 50(3), 533–603.
- Lerner, G. H. (1991). On the syntax of sentences-in-progress. *Language in Society*, 20(3), 441–458.
- Sacks, H. (1995). *Lectures on conversation: Volume I (1964–1968)*. (G. Jefferson, ed.) Blackwell.
- Slobin, D. I. (1991). Learning to think for speaking. *Pragmatics*, 1(1), 7–25.
- Speck, A. (1995). *Textproduktion im Dialog. Zum Einfluß des Redepartners auf die Textorganisation*. Westdeutscher Verlag.
- Stalnaker, R. C. (2002). Common ground. *Linguistics & Philosophy*, 25, 701–721.
- von Stutterheim, C. (1997a). Zum Ausdruck von Raum- und Zeitkonzepten in deutschen und englischen Texten. *Zeitschrift für Germanistische Linguistik*, 25, 147–166.
- von Stutterheim, C. (1997b). *Einige Prinzipien des Textaufbaus: Empirische Untersuchungen Zur Produktion Mündlicher Texte*. De Gruyter.
- von Stutterheim, C. (2001). Hörerorientierung in der Kommunikation. Psycholinguistische Evidenz pro und contra. In A. Lehr, M. Kammerer, K.-P. Konerding, A. Storrer, C. Thimm, & W. Wolski (Eds.), *Sprache im Alltag. Herbert Ernst Wiegand zum 65. Geburtstag gewidmet* (pp. 467–484). De Gruyter.
- von Stutterheim, C., & Kohlmann, U. (1998). Selective Hearer Adaption. *Linguistics*, 36(3), 517–549.
- von Stutterheim, C., Carroll, M. (1993). Raumkonzepte in Produktionsprozessen. *Kognitionswissenschaft*, 3(2), 70–82.
- von Stutterheim, C., & Carroll, M. (2018). Texts as answers to questions. In A. Hübl, & M. Steinbach (Eds.), *Linguistic Foundations of Narration in Spoken and Sign Languages* (pp. 67–92). John Benjamins.
- Watorek, M. (2003). The development of anaphoric means to refer to space and entities in the acquisition of French by Polish learners. In C. Dimroth, & M. Starren (Eds.), *Information Structure and the Dynamics of Language Acquisition* (pp. 329–356). John Benjamins.

## Appendix

### Stimulus pictures



Stimulus picture 1



Stimulus picture 2

---

Silvia Natale

# The Use of Progressive Aspect in Two Varieties of Swiss German

---

**Abstract** The present study explores the use of the progressive *am*-periphrasis in Swiss High German and in the Bernese dialect, since the progressive form is considered particularly vital in both Swiss High German and Swiss German dialects. Based on a scene description task, data were collected at a Bernese high school, allowing to determine differences in the frequency of use of the *am*-periphrasis. Furthermore, it is analyzed which semantic attractors of the items used in the video affect the selection of the *am*-periphrasis. Another analysis concerns the verbal representation of the events (macro- or micro-event) in order to highlight connections between the selection of form and temporal conceptualization. The argument structure of the verbs used was also considered to identify differences with regard to the incorporation of direct objects.

**Keywords** progressive aspect; Swiss German; conceptualization of events.

## Introduction

All natural languages can express that an event being referred to is “ongoing at a reference time” (Bybee et al., 1994, p. 126), regardless of whether they have grammatical or lexical means to convey a progressive view of the event.<sup>1</sup> German, as well as the Swiss German dialects, dispose of the verbal periphrasis *am* + infinitive + *sein* (henceforth *am*-periphrasis), which expresses if *sein* is anchored in present tense, that an event is anchored in the *here and now* and can be assumed to continue in the immediate future (Bertinetto, 1986).

Longstanding typological studies have shown (Dahl, 1985; Bybee et al., 1994) that the development of means to express progression can follow a

---

1 The present study was planned and carried out in honor of Christiane von Steutterheim. When I moved to Bern, she inquired about the use of the *am*-periphrasis in German-speaking Switzerland. After many years of providing subjective information, this study may better satisfy her interest in this field.

number of paths. In this context, *obligatorization* proves to be an important parameter in the study of processes leading to the grammaticalization of the relevant forms of expression (Bybee & Dahl, 1989; Lehmann, 2015). This makes Swiss High German and Swiss German dialects, the languages in question, an interesting and relevant object of continued research<sup>2</sup> since the use of verbal periphrases to indicate progression is not yet mandatory and does not exclude, for example, a habitual or generic reading.<sup>3</sup>

In Standard German, as well as in Swiss German dialects, the verbal *am*-periphrasis expresses progressivity, but use is not obligatory. Since Swiss German dialects differ significantly from Standard German with respect to the frequency of use of progressive forms (Flick & Kuhmichel, 2013), a systematic comparison of the contexts of use for both varieties is relevant. In the present pilot study, which has been carried out in Bern/Switzerland, focus is placed on the specific forms that the study participants use spontaneously when describing ongoing events presented in short video scenes (see details in chapter 3). Given the specific range of events, the study participants' choice of expression provides information on relevant features of the progressive form compared to its counterpart (verb in the present indicative) which is underspecified in this respect. The first question that emerges is thus:

Are there differences with regard to both the frequency as well as the contexts of use for the *am*-periphrasis in the Swiss German dialect (here and henceforth Bernese) compared to Standard German (in this case, Swiss High German)?

Another question concerns semantic attractors that may influence the selection of the *am*-periphrasis. The semantics of progressive verbal periphrases can be determined in terms of their aspectual function, which can change in the course of grammaticalization (e.g., from locativity over durativity to progressivity, cf. Comrie, 1976; Bybee & Dahl, 1989; Bybee et al., 1994; Bertinetto et al., 2000; Killie, 2008). Functional development is closely connected to compatibility with different verb types. This can vary depending on the degree of grammaticalization. Thus, at the beginning of the grammaticalization process (Bybee et al., 1994), activity verbs increasingly occur in constructions that, in the sense of Vendler (1967), describe a dynamic, durative, or atelic situation. In the course of grammaticalization, the semantic radius expands and allows

---

2 If use is not obligatory, a change in compatibility with verb classes (for example, the extension to static verbs) or shifting functions (for example, durative or focalizing, cf. Bertinetto, 2000) takes place.

3 This is the case in English, for example, where, with a few exceptions, use of progressive aspect is obligatory in certain contexts, cf. Mair, 2012).

for combination with telic verbs. At an advanced stage of grammaticalization, static verbs can also combine with progressive constructions, but this also marks the transition from progressivity to pure imperfection (Bybee et al., 1994). The property of progressive aspect in defocusing the boundaries of an event promotes its occurrence with activity verbs where the absence of an intrinsic boundary is a typical feature (Ramat, 1995, pp. 44–45; Sasse, 2002, p. 206). The semantic compatibility of activities with the *am*-periphrasis, or the existing affinity between the *am*-periphrasis and activity verbs, has been investigated in “corpus-linguistic” studies. Flick and Kuhmichel (2013), for example, could demonstrate for press language in German that the compatibility of the *am*-periphrasis and activity verbs is the highest. Kempf (2023) also found for Swiss German that the preference for activity verbs is most pronounced in the Swiss SMS corpus. The question that therefore arises is:

Are there potential differences regarding the event type that may affect the selection of the *am*-periphrasis?

Another question is linked to the verbal representation of the events. In the present study, dynamic events in progress were used in the form of video clips in order to quantify the choice of progressive constructions and to determine frequency of use with respect to specific situation types. In addition to these quantitative and semantic questions, the present framework can also be used to provide insights into language-specific patterns in event conceptualization. The way an event is conceptualized (cf. Levelt, 1989, 1999) reveals patterns with regard to the selection of information, its decomposition and linearization (cf. von Stutterheim & Nüse, 2003). Various comparative language studies have shown (cf. Carroll & von Stutterheim, 2010; von Stutterheim et al., 2017) that there is a link between language-specific categories which are grammaticalized and specific patterns in the conceptualization of events. One grammatical factor that has been shown to be crucial in this regard is the presence of the grammatical feature “progressive aspect” as encoded on the verb (Carroll & Lambert, 2003; 2006; von Stutterheim & Lambert, 2005; Carroll et al., 2008; Natale, 2018). Dynamic events can show different levels of granularity in their verbal representation, as expressed via verb semantic categories. An everyday scene shown in an experiment in which a child catches and throws a ball can be represented in its entirety, for example, by the utterance “A child is playing with a ball”. A holistic mode of representation of this kind, which conceptualizes an event as a macro-event, can be contrasted with representations that relate to its micro-events, as in the example sentence “The child catches the ball and then throws it away again”. A study by Carroll and von Stutterheim (2010) showed how German and English differ with respect to event representation. One relevant factor

can be attributed to the presence of the -ing form, which represents an event as ongoing, in contrast to the simple tense (*She is playing tennis* versus *She plays tennis*). The question that therefore arises is:

How do informants represent the events verbally?

## 1 The *am*-periphrasis in German and Swiss German

The relevant means to express progression in both Standard German and Swiss German dialects is the *am*-periphrasis. The *am*-periphrasis in Standard and Swiss is structured as follows: *am* + infinitive + *sein* (*Er ist am Arbeiten*).

The *am*-periphrasis represents an event as “in progress” or as “not completed” (Flick & Kuhmichel, 2013). The aspectual distinction underlying this function is considered a special category of the imperfective aspect. In Klein’s (1994) terminology, the progressive aspect is characterized by the fact that the assertion time, or topic time (TT), i.e., the period of time about which an assertion is made in the predication, is contained within the situation time (TSit), the period of time for which a situation holds. The often-cited viewpoint that the imperfective aspect presents an event partially, or from the “inside” (cf. Comrie, 1976; Bertinetto, 1986), is plausible with the inclusion of TT in TSit.

In the research literature on German, the *am*-periphrasis has received increasing attention in recent years (cf. Ebert, 1996; Krause, 1997 and 2002; Glück, 2001; Flick & Kumichel, 2013; Gárgyán, 2014; Ferraresi, 2017). A detailed description of the increasing grammaticalization of the form as mentioned in the literature (as on its way to becoming “a grammatical sign”, cf. Kempf, 2023, p. 3) will be omitted here. Reference is made to Kempf (2023), who traces the status of the progressive *am*-periphrasis in Swiss German diachronically and analyses its status in the Swiss German verbal system on the basis of various data sets. The relevant features in the context of the present study are as follows:

The *am*-periphrasis is not obligatory and is, therefore, (still) optional with regard to the parameter of obligatoriness, which is relevant for grammaticalization (cf. Lehmann, 1995 and 2005). The *am*-periphrasis has two regional core areas, which are located both in the Rhineland-Westphalia region (Brons-Albert, 1984; Ramelli, 2015a and 2015b) as well as in German-speaking Switzerland (cf. van Pottelberge, 2004; Ebert, 2000; Flick & Kuhmichel, 2013). The descriptions of the *am*-periphrasis available so far mainly refer to newspaper texts (cf. Flick, 2016), corpora of written (van Pottelberge, 2004) as well as spoken language (cf. Krause, 2002) or questionnaires (Reimann, 1997; Kempf, 2023).

## 2 Method

### 2.1 General remarks

In order to provide an answer to the research questions in section 1, we will draw on a well-established empirical method that originated in Christiane von Stutterheim’s DFG-project “Konzeptualisierung und einzelsprachliches Wissen in der Sprachproduktion”<sup>4</sup>. This has been applied in a number of cross-linguistic studies on the use of forms to express viewpoint aspect (cf. von Stutterheim & Nüse, 2003; Carroll et al., 2008; Natale, 2009; Flecken, 2011).

The experiment is based on 30 short video clips that depict a range of situations with different properties (see 2.2 for a detailed description of the items). Given this experimental approach, control variables can be determined that can reveal the role of semantic properties depicted in the video clips with the selection of a viewpoint “event is in progression” and the corresponding verbal periphrases. It is thus possible to isolate the properties that promote the use of progressive verbal periphrases. The contrastive study also benefits from this procedure, as the comparability of the data is guaranteed, and the use of the means to express the viewpoint “event in progression” in the two varieties studied can be compared on a systematic basis.

The use of this experimental design in comparing progressive verbal periphrases in Swiss High German and Bernese presents an additional approach to studies based on the analysis of corpora as well as questionnaires in translation tasks or acceptability assessments.

### 2.2 Design of the scene description task

The scene description task consists of 30 video clips with durations of 5 and 8 seconds in which all events are depicted as ongoing (e.g., *a person swimming in a lake* or *two men playing chess*). All the videos shown in the current study can, therefore, potentially trigger the selection of the *am*-periphrasis.

Table 1 lists the situations shown in the video and indicates to which semantic category they belong. A blank screen was shown in between each video clip. Figure 1 shows an example of a video clip.

---

4 The project with the number 5381325 was funded by the German Research Foundation (DFG) from 1997 to 2004.

Table 1. Overview of the stimuli used

No.	Content of the scene
1.	Sandcastle: Two girls are building a sandcastle on the beach and are decorating it.
2.	Engine: A man is working on the engine of a car.
3.	Studying: A woman is sitting at a desk. She is reading and underlining something in a text.
4.	Shower: A man is standing under the shower shampooing his hair.
5.	Toothpaste: A man is applying toothpaste to his toothbrush.
6.	Playing ball: A child is in a living room and is catching a ball.
7.	Coffee: A person is standing at a coffee machine and is filling a filter with coffee.
8.	Baking: Ingredients such as butter and sugar are shown being stirred together to make a dough.
9.	Planting: Two people are planting a plant.
10.	Hairdresser: A hairdresser is applying color to someone's hair with a brush.
11.	Sprint: A person is getting ready for a sprint at the starting block.
12.	Typewriter: A man is sitting at a desk and is putting a sheet of paper into a typewriter.
13.	Formula One: A man is walking to a Formula One car and sits in the car.
14.	Watch: A woman is adjusting her fitness watch.
15.	Plane: An aircraft is landing on a runway.
16.	Pizza: A woman is eating a piece of pizza.
17.	Ironing: A man is ironing a shirt together with his son.
18.	Chess: Two men are playing chess.
19.	Key: Someone is opening a door with a key.
20.	Wine: A man is drinking a glass of wine.
21.	Bike: A man is riding a bike.
22.	Swimming: A person is swimming in the open sea.
23.	Reading: A woman is sitting in an armchair reading a book.
24.	Guitar: Hands are shown playing the guitar.
25.	Tennis: A man is playing tennis in a hall.
26.	Crocheting: Someone is crocheting.
27.	Drawing: Someone is drawing a face.
28.	Cup: Someone is making a cup out of clay.
29.	Snowman: Two children are building a snowman.
30.	Tailor: A man is pinning a dress on a bust.





Figure 1. Screenshot of video number 29 (*snowman*)

With regard to possible differences in the area of attractor effects, the clips belong to different situation types<sup>5</sup> which vary with regard to their semantic and temporal properties. Amongst others (which are not relevant here as they were used as fillers), causative events (such as “making a mug” or “knitting a scarf”) and activities (such as *riding a bicycle* or *playing the guitar*) were distinguished.

Activities are dynamic and durative situations and are characterized by the homogeneity of their sub-events (such as the regular swimming strokes in swimming). Each individual event in the process can be representative of the whole event. Situations of this kind are also termed *atelic* in the literature (cf. Ryle, 1949; Vendler, 1967; Comrie, 1976; Smith, 1991; Sasse, 2002) because they do not have an inherent endpoint. As stated above, in the present study, activities are contrasted with causative events (for example, *building a snowman* or *pottering a mug*). Causative events, unlike activities, are characterized by having a potential intrinsic endpoint (when the mug is finished being potted), which implies a change of state. According to the terminology in Klein (1994), this provides a contrast between an initial state and a final state. Given the relationship between progressive marking and activities as established in the research literature, the following analysis investigates to what extent activities trigger higher frequencies in the use of progressive verbal periphrases.

---

5 By situation types we do not mean here the temporal properties of verbs or of idealized situations associated with verbs (cf. Smith, 1991), but rather the temporal semantic properties of an extra-linguistic situation contained in the videos.

A further question deals with the semantic features of events which are closely related to aspect marking. The question arises as to which observations can be made with regard to the level of granularity in the linguistic representation of the events (cf. Carroll, Rossdeutscher, et al., 2008). Events can be represented as a whole or as a sub-event of a superordinate event. In order to investigate whether the speakers represent an event as a macro-event (as a whole) or as a micro-event (as a sub-event of a higher-level event), some situations that potentially allow both perspectives on the event have been inserted in the design for a preliminary analysis. They include *making coffee* (micro event: *someone is filling a filter with ground coffee*; macro event: *someone is making coffee*) or *learning* (micro event: *someone is reading and underlining something in a text*; macro event: *someone learning/working*).

With regard to the question of their syntactic realization, situations were chosen that allow for different representations in this respect. In scenes in which “a woman is eating a slice of pizza”, different utterances such as *she eats/is eating* (unbounded event) or *she eats (a slice) of pizza/she is eating pizza* are conceivable. Other scenes used for the analysis of the argument structure are, for example, *baking a cake* (*she is baking* vs. *she is baking a cake*) or *crocheting a scarf* (*she is crocheting* vs. *she is crocheting a scarf*).

### 2.3 Overview of the informants

The scene description task was carried out with 40 native speakers of Swiss German who were randomly assigned to two groups. One carried out the scene description task in Swiss High German and one in Bernese dialect. The informants are students at a high school in Bern (Gymnasium) and were aged between 15 and 18 (28 male and 12 female).

### 2.4 Experimental procedure and data processing

The participants took part in the scene description task individually and were given the following instruction: “You will now see 30 short videos on the computer in which different people perform an action. This will take 5 minutes. We will now watch the video for each scene and you will be asked ‘Was passiert hier?’. You do not have to mention details, like the clothes people are wearing, etc. Just say ‘what happens.’”

The responses were recorded, transcribed, and analyzed using the MAX QDA programme.

### 3 Results<sup>6</sup>

#### 3.1 Quantitative analysis

##### 3.1.1 Overview

The data analysis showed that there is a wide spectrum in the variety of forms in both Swiss High German and Bernese. An overview of the selected verb forms is listed in table 2, with an example in each case from the corpus. Differences with regard to the explicit use of the relevant forms became apparent in both data sets and are listed in the following chapter.

Table 2. Text examples from the corpus

Verbal form	Swiss High German	Bernese dialect
Present Indicative	Ein junger Mann trinkt ein Glas Weisswein (D_Vp21_Juri, Pos. 9)	E Mah trinkt Wii. (CH_Vp13_Giovanna, Pos. 9)
<i>am</i> -periphrasis	Jemand ist sich am Haare waschen (D_Vp31_Amina, Pos. 18)	Es isch e Person am Bache (CH_Vp16_Jan, Pos. 32)
<i>tun</i> + infinitive	Ein Vater mit dem Kind und die tun das Hemd bügeln (D_Vp36_Maria, Pos. 24)	E Vater tuet mit sim Sohn es Hemdli bügle (CH_Vp1_Anna, Pos. 27)
Reduplication with <i>gang/go; chum/cho</i>	–	Öpe- e Frou, wo im Waud geit go Jogge. (CH_Vp9_Yann, Pos. 74) öper wo gad mitemne Heum chunnt cho z loufe (CH_Vp9_Yann, Pos. 43)
<i>Dabei sein</i>	Man sieht eine Person beim Kraulen im See, Meer, irgendetwas (D_Vp27_Valentina, Pos. 53)	E Frou i Sportchleider isch äüä derbii z go jogge (CH_Vp1_Anna, Pos. 67)

6 With 40 participants and 30 items, the sample does not allow a statistically relevant analysis. The following results should therefore be treated as trends.

Table 2. (continued)

Verbal form	Swiss High German	Bernese dialect
Periphrasis with <i>gehen</i>	Ich sehe eine Person, die jetzt dann wahrscheinlich Sprinten oder Joggen geht (D_Vp22_Luana, Pos. 14–15)	–
Present participles	Startendes Flugzög (CH_Vp2_Livia, Pos. 63)	–
Infinitive	Wahrscheinlich in einer Autowerkstatt, den Motor ausbauen, Oder reparieren (D_Vp28_Emma, Pos. 9–10)	Schribmaschine schribe... äüä amne Text (CH_Vp2_Livia, Pos. 19)
Noun	Ich sehe ein Formel Eins Auto (D_Vp24_Carmen, Pos. 42)	Wahrscheinlech es Usdur- oder Sprinttraining ufem rote Platz. (CH_Vp19_Timo, Pos. 13)
Future	Jemand, der sich die Zähne putzen wird (D_Vp36_Maria, Pos. 19)	–
Passive	Es wird eine Sandburg gebaut (D_Vp25_Anouk, Pos. 5)	Es wird es Gsicht zeichnet, mit Orange (CH_Vp12_Fabiana, Pos. 42)

### 3.1.2 Quantitative analysis in Swiss High German and Bernese

In Swiss High German the preferred form is the simple form in the present indicative (73.7 percent), as table 3 shows. This is followed at a considerably lower frequency of the *am*-periphrasis (at) 16.0 percent. The passive was used in 31 cases (5.2 percent). The remaining forms occur at low values over a broad spectrum but do not show any significant frequencies. In another study by Flecken (2011, p. 206) in which data were collected in Germany on the basis of a comparable framework, the frequency for the *am*-periphrasis in High German amounts to 1.41 percent (22/1556), showing a marked quantitative contrast to the data recorded in Switzerland.

Table 3. Overview of the chosen verb forms and their frequency in Swiss High German and Bernese

Verbal form	Swiss High German (n= 600)	Bernese Dialect (n= 600)
Present indicative	73.7	49.5
<i>am</i> -progressive	16.0	28.2
<i>tun</i> + infinitive	0.2	11.3
Reduplication of the type <i>gang/go, chum/cho</i>	0	0.8
<i>Dabei sein</i>	0.5	0.2
Periphrasis with <i>gehen</i>	0.3	0
Present participle	0	0.2
Infinitive	0.3	4.8
Noun	1.5	1.5
Future tense	0.7	0
Passive	5.2	1.3

As in Swiss High German, the present indicative is the preferred form in Bernese. However, there is a considerably higher value for the *am*-periphrasis, which at 28.2 percent is almost twice as high as the data for Swiss High German (16.0 percent). This was to be expected given the higher occurrences of the *am*-periphrasis in the research literature for Swiss German dialects. Likewise, the relatively frequent use of *tun*-periphrasis in Bernese is consistent with the descriptions in the literature (cf. Krause 2002; van Pottelberge, 2004). As in Swiss High German, the other forms show a low frequency. The relatively high value for infinitive constructions (4.8 percent), however, is due to the fact that one speaker produced 19 of the 26 utterances in the infinitive.

### 3.1.3 Temporal adverbs

The frequency of occurrence for temporal adverbs such as *jetzt* or *gerade* was evaluated for both varieties. Table 4 shows that the simple form of the present indicative is frequently combined with temporal adverbs (in 73 percent of the cases in Bernese and 86 percent in Swiss High German). This high value acts in favor of the spreading progressive verb forms as temporal adverbs produce the aspectual reading lacking in the simple form.

Table 4. Overview of the use of temporal adverbs in Swiss German and High German in connection with chosen verb forms

	<i>am</i> -periphrasis	Present	<i>tun</i> -periphrasis
Bernese	24 % (27)	73 % (82)	4 % (4)
Swiss High German	14 % (8)	86 % (49)	0 % (0)

We can state that in the present scene description task in which the participants could spontaneously and freely select the verb form, the simple form of the present indicative is preferred in both varieties. The more advanced grammaticalization of the *am*-periphrasis in Swiss German dialects, as discussed in the research literature (cf. Kempf, 2023), is reflected in a percentage that is twice as high in Bernese compared to the data for Swiss High German (28.2 percent vs. 16.0 percent). In comparison to English, however, which is considered a prime example of a fully grammaticalized progressive, these values show that in contexts where use would be obligatory, Bernese is still far from a stage of grammaticalization that correlates with very high frequencies of use. A comparison with Italian, in which the use of the progressive verbal periphrasis *stare + gerundio* is steadily increasing, shows that the progressive verbal periphrasis *stare + gerundio* in Italian is used more frequently.<sup>7</sup>

### 3.1.4 Individual differences

A more detailed analysis of the frequency of use shows that in both the Swiss High German and the Bernese data, the choice of the *am*-periphrasis varies from person to person (see table 5). Figure 2 (absolute values per participant, potential maximum value = 30) depicts this heterogeneity by means of the oscillating curve.

Table 5. Swiss High German group

Subjects and frequency of the <i>am</i> -periphrasis																			
01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
5	2	0	2	0	1	3	7	0	0	2	9	5	19	0	5	0	14	7	8

<sup>7</sup> In an analogous experiment, a value of just under 42 percent is achieved in Italian for the verbal periphrasis *stare + gerundio* (cf. Natale, 2019; Natale, 2009), showing a more frequent use of *stare + gerundio*.

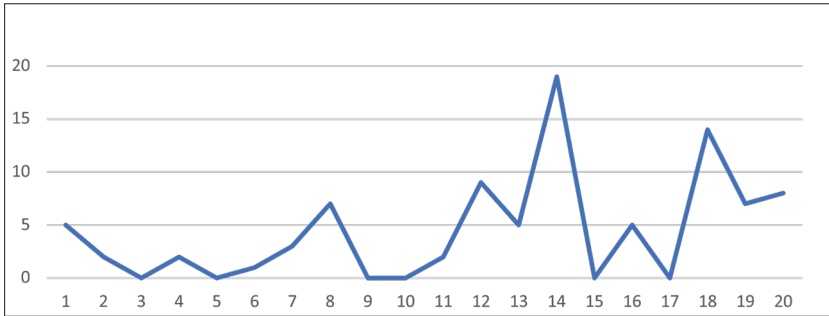


Figure 2. Variability progressive by subject (absolute values)

When verbalizing the events shown in the video (expressed in figure 3 as a percentage), we see that one participant selects the *am*-periphrasis at a rate of 63 percent, while 6 participants do not select it at all, however. The rate for 5 participants is below 10 percent and for a further 7 it varied between 10 percent and 30 percent. One participant, however, used the *am*-periphrasis in almost half of the cases (47 percent).

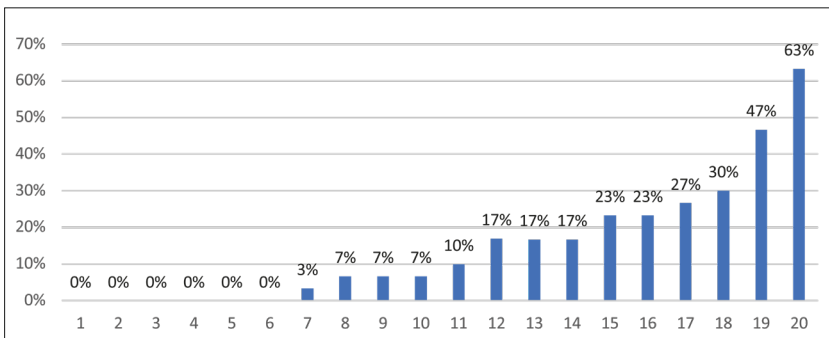


Figure 3. Individual variation group High German (%) ordered by frequency

In the Bernese data, an individual variation can also be observed (see table 6), which is also expressed in the oscillating curve of figure 4 and the ordered frequencies in figure 5.

Table 6. Test person group Bernese.

Subjects and frequency of the <i>am</i> -periphrasis																			
01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
12	0	10	0	0	3	16	5	2	23	21	0	1	1	9	14	2	20	17	13

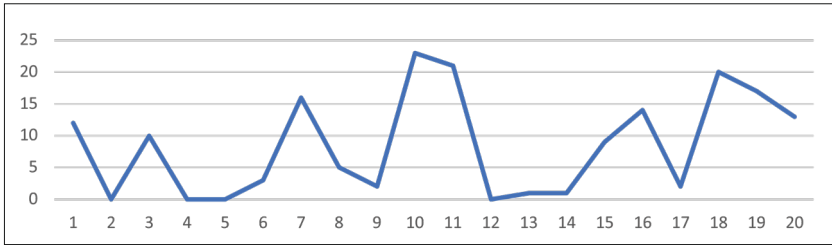


Figure 4. Swiss German: progressive choice by subject

Corresponding to the higher frequency of the *am*-periphrasis in Bernese, the values are higher overall. However, 4 participants in this group do not use the *am*-periphrasis at all, while 3 test subjects use it relatively frequently with over 60 percent.

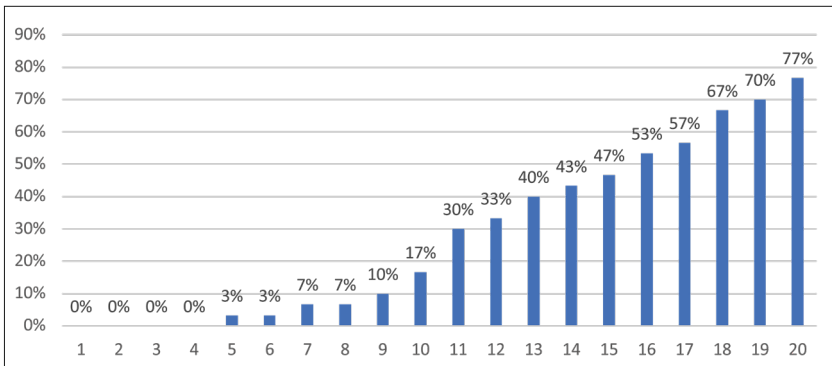


Figure 5. Swiss German: progressive choice by frequency

## 3.2 Semantic attractors

### 3.2.1 Cross-linguistic studies on factors determining use of the progressive

The use of the progressive when describing ongoing situations presented in video clips (cf. von Stutterheim & Nüse, 2003; Gerwien & von Stutterheim, 2022a; 2022b) revealed the interplay between specific semantic properties of situation types and their relevance for aspect marking. In the following, we refer to three studies which, based on this form of stimulus material, set out to investigate the factors that lead to the use of constructions that mark an event as ongoing / in progression in French (cf. Leclercq, 2008), Italian (cf. Natale, 2009; 2019) as well as Dutch and Standard German (cf. Flecken, 2011).



The languages in question, Italian, French, and Dutch, as well as Swiss High German and Bernese all have constructions which can mark an event as ongoing/in progression construction.<sup>8</sup> It should be stated, however, that use is not obligatory in these languages. In this sense, the constructions in question are described as emergent in the literature (cf. for Italian Bertinetto, 1990; Berretta, 1993; Strudsholm, 2004; Berruto, 2017; for French Pusch, 2003; 2005; for Dutch Flecken, 2011).

Studies by Leclercq (2008), Natale (2009), and Flecken (2011) examine, among other factors, the relationship between activities and causative events, in addition to further situation types (motion events, static situations) which are not discussed here. However, before discussing the role of the situation types in the selection of a form that expresses the meaning “*event is ongoing/in progression*”, differences in the frequency of the relevant verbal periphrases should be considered, as illustrated in figure 6.

In the two Romance languages, Italian and French, there is a clear quantitative difference given the fact that in Italian the frequency of the verbal periphrasis *stare + gerundio* is almost twice as high (41.95 percent) as in French (22.10 percent). In the two Germanic languages, there is a clear difference since the frequency of use in Dutch of the relevant form *aan het + infinitive* at 27.52 percent is significantly higher compared to the *am*-periphrasis in the standard German data (1.41 percent).<sup>9</sup>

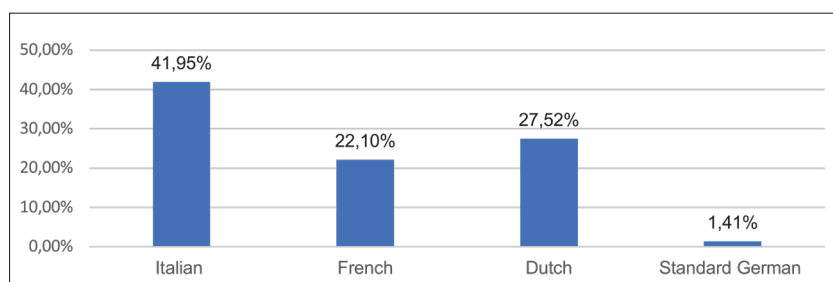


Figure 6. Use of the progressive verbal periphrasis in comparable situations

Given the contrasts in the “attraction” effect of the stimuli used (e.g., activities or causative events) it could be shown in Natale (2009; 2018) that, as in French (cf. Leclercq, 2008), compared to activities, causative events are more likely to be verbalized using progressive verbal periphrases. In Natale (2009), it is argued that the temporal contrast between the different states of an event, as

8 In Italian the verbal periphrasis is *stare + gerundio*; in French *aller + infinitive* and in Dutch *aan het + infinitive*.

9 On the data recorded in Germany (Heidelberg) see chapter 3.1.2.

expressed by causatives, forms a core concept of progression and therefore evokes a higher use of progressive constructions in the initial stages of development. The lack of substantial results in activities, which, on the other hand, is inherent to causative events, correlates with a lower use of *stare + gerundio*. In Dutch, on the other hand, both activities (event is ongoing), as well as causative events, attract the use of the periphrasis *aan het + infinitive to the* same extent, thus indicating no preference in this respect.

### 3.2.2 Attractors in Swiss High German and Bernese

In the present stimulus material, the affinity between progressivity and activity verbs was tested via semantic attractors by varying a set of different situation types. Now we come back to the results.

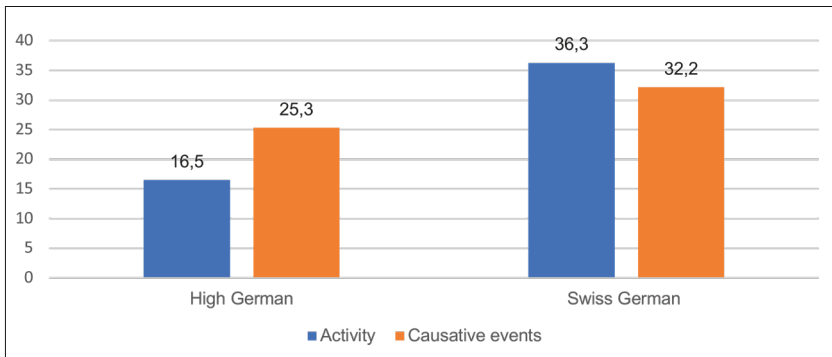


Figure 7. Activities and causative events in Swiss High German and Bernese

Figure 7 shows that the Swiss High German data differ from the Bernese data with regard to the rate of attraction of the situation types. While the Swiss High German data show higher use of the *am*-periphrasis with causative events compared to activities (25.3 percent in contrast to 16.5 percent for activities), the Bernese data does not show a clear preference. The difference in the use of the *am*-periphrasis between activities at 36.3 percent compared to 32.2 percent for causative events is minimal. The preference observed in Swiss High German and the two Romance languages is not evident in Bernese. The data for Bernese reflect a tendency that has also been observed in Dutch, as mentioned above (Flecken, 2011; Behrens et al., 2013). The fact that situations with a change in state (causative events) as well as situations without a change in state change (activities) both trigger a high use of progressive constructions was attributed for Dutch to the role of homogeneity for the represented events. This factor was present in both situation types. Whether this affinity is also present in Bernese would have to be re-examined

in a subsequent study with a more refined design. It should be noted at this point that the Swiss High German and Bernese data differ with regard to the effect of the situation types and may therefore have taken different paths in the process of grammaticalization.

The question of why Swiss High German shows, like Romance languages, a preference for causative events while Bernese does not may be answered with a sensitivity to homogeneity that goes hand in hand with a different level of grammaticalization.

### 3.3 Verbal representation of the events

In the present study, six items were examined in a first step in the analysis with regard to their verbal representation. The items were chosen as they allow for either a representation as macro- or as micro-event. Are there preferences with regard to the representation of an event as a macro- or micro-event? The overview in table 7 illustrates the depicted scenes in more detail and, using examples from the corpus, the linguistic means that were selected in response to the specific scene.

Table 7. Examples of event representation in Swiss High German

Item	Description	Macro-event	Micro-event
Studying	A woman is sitting at a desk. Reads a text, underlines and turns the page.	Ich sehe eine Frau, welche am Lernen ist (D_Vp24_Carmen, Pos. 20)	Eine Frau beugt sich über einen Ordner Und schreibt etwas auf auf ein Blatt. (D_Vp37_Miriam, Pos. 15–16)
Shower	A man in the shower and shampooing his hair.	Ein Mann ist am Duschen, (D_Vp33_Laura, Pos. 20)	Ein Mann, der sich die Haare wäscht, Oder einschamponiert, unter der Dusche. (D_Vp32_Janina, Pos. 18–19)
Playing ball	A child stands in the living room, catches a ball and throws it away.	Ein kleiner Junge, der Torwart spielt. (D_Vp28_Emma, Pos. 29)	Ein Kind versucht, einen Ball zu fangen. (D_Vp27_Valentina, Pos. 30)
Making coffee	A man stands at a coffee machine and fills the hand sieve with ground coffee.	Ich sehe eine Person, die Kaffee macht. (D_Vp30_Angela, Pos. 35–36)	Kaffeepulver wird zusammengedrückt und dann in die Kaffeemaschine ... (D_Vp35_Hanna, Pos. 38)

Table 7. (continued)

Item	Description	Macro-event	Micro-event
Baking	Ingredients such as butter, sugar and flour are shown being mixed together.	Hier ist jemand am Backen (D_Vp26_Angela, Pos. 35)	Jemand, der gerade Zutaten zusammenmischt. (D_Vp36_Maria, Pos. 30)
Tailor	A man stands in front of a dress bust and pins a dress.	Ein Designer, der seine Kleidung näht. (D_Vp28_Emma, Pos. 69)	Jemand steckt Nadeln in eine Puppe, ja. (D_Vp29_Saphira, Pos. 68)

The second step in the analysis asked to what extent the degree of granularity (macro-event vs. micro-event), as expressed in the verb, goes hand in hand with the presence of an aspectual distinction.

### 3.3.1 Representation as macro-events or as micro-events

Both Swiss High German and Bernese show a preference for representing the selected items as macro-events (67 percent for macro-events compared to 33 percent for micro-events in Bernese; in Swiss High German the preference for macro-events is more pronounced with 74 percent compared to 26 percent micro-events [see table 8]).

Table 8. Overview of macro and micro events in Swiss High German and Bernese

	Macro-event	Micro-event
Bernese (n= 86)	67% (58)	33% (28)
Swiss High German (n= 105)	74% (78)	26% (27)

While at first sight there is no marked difference between the two varieties with respect to macro- and micro-events, figure 8 shows that in Bernese macro-events are coded more often with the simple form of the present indicative (53.4 percent) than with the *am*-periphrasis, although it nevertheless shows a relatively high value at 46.6 percent. In Swiss High German, on the other hand, there is a clear preference to code macro-events using the simple form of the present indicative.

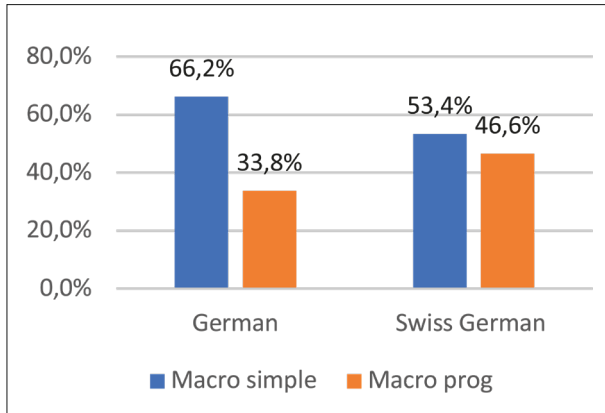


Figure 8. Macro-event and *am*-periphrasis in Bernese and Swiss High German

Given the representation of the event at a low level of detail (see figure 9), the *am*-periphrasis is used less frequently in both varieties, although in Swiss High German the contrast between the simple form (89.2 percent) and the *am*-periphrasis (10.8 percent) is more marked than in Bernese (67.9 percent simple form present indicative versus 32.1 percent *am*-periphrasis). The fact that the affinity toward the *am*-periphrasis is higher for macro-events than micro-events in both varieties may be explained by its property in defocusing (Langacker, 2000) or fading out (Ebert, 1996) event boundaries.

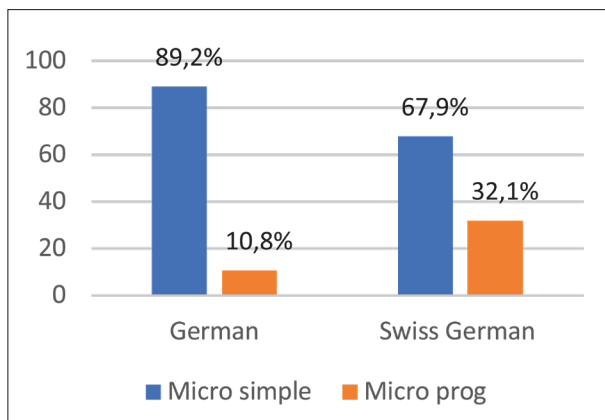


Figure 9. Event representation in Swiss High German

The analysis of the chosen levels of representation represents only one factor in the field of temporal conceptualization. In the next step, we will examine whether the preference for the macro-level of representation with the defocusing of boundaries is also reflected in the syntactic realization.

### 3.3.2 Syntactic realization

In the analysis of the syntactic realization five items (*crocheting, baking, dressmaking, painting, making a mug*) were selected. These items allow the naming of an object, but the naming of the object is not indispensable, as for example in the case of the item *crochet*. The statement *someone is crocheting* could be just as acceptable as *someone is crocheting a scarf*. The present analysis looks (i) at the frequency with which direct objects are mentioned and (ii) how the objects are syntactically embedded. In both German and Swiss German dialects the syntactic embedding allows conclusions to be drawn on the degree of grammaticalization for the use, for example, of the *am*-periphrasis with direct objects (cf. also Kempf, 2023, and Ramelli, 2015b). In three of the five items, the effected object must be inferred because it is not visible in the scene. This may affect the mention of the syntactic object with use of a definite article or not.

- 1) To crochet (x)
- 2) To bake (x)
- 3) To dressmake (x)

The effected object is clearly visible in two items:

- 1) Painting a picture
- 2) Making a mug

The evaluation was confined to the statements in which the verb occurs and thus allows mention of a direct object (bivalent verbs). This means that the first statement of the informant was evaluated in each case but not the subsequent utterances. For example, if a speaker responds to the scene with *Someone is baking. Maybe a cake is being made*, the unit analyzed was confined to the first utterance as this can be considered more relevant with respect to the overall conceptualization of the event.

Even though the data set is limited, we can nevertheless observe a relevant tendency as indicated in table 9: Objects are mentioned, albeit with a slightly higher frequency in the data for Swiss High German compared to the Bernese data (57 percent in High German versus 42 percent in Swiss High German).

Table 9. Frequency of direct objects in comparison

	Object mentioned	Object not named
Swiss High German	51/89 = 57 %	38/89 = 43 %
Bernese	37/88 = 42 %	51/88 = 58 %

Use of the progressive is accompanied by a lower frequency in the mention of direct objects. The relevant feature of progressive in the present context (event is encoded as *ongoing*), which may also defocus the boundaries of an event, is accompanied by a lower frequency in the mention of direct objects as shown by the following results: In both varieties, the *am*-periphrasis is more likely to be used when the direct object is not mentioned.

When a direct object is mentioned in the data for High German, the frequency of use of the *am*-periphrasis is 22.2 percent. If the direct object is not mentioned, the frequency of the *am*-periphrasis increases to 77.8 percent. This corresponds to the situation in Swiss German. If a direct object is mentioned the occurrence of the *am*-periphrasis is 21.6 percent. If it is not mentioned, the frequency increases to 78.4 percent. In both varieties, there is therefore a very clear tendency for the naming of the direct object to be accompanied by a reduction in the use of progressive constructions.

With regard to the syntactic embedding, we see for the scene description task that reference to the direct object varied as shown in table 10 with examples from the corpus.

Table 10. Overview of the realization of the direct objects

	Swiss High German	Bernese
“something/etwas”	Eine Frau, die etwas strickt, etwas Graues. (D_Vp23_Luana, Pos. 47)	Me gseht Häng wo öpis töpfere (CH_Vp12_Fabiana, Pos. 61)
Indefinit Sg.	Ich sehe eine Schüssel und eine Person, die einen Kuchen oder so etwas ähnliches backt. (D_Vp24_Carmen, Pos. 39–40)	Es wird es Gsicht zeichnet, mit Orange, (CH_Vp12_Fabiana, Pos. 42)
Plural without articles	Ich sehe einen Mann, der Kleider näht. (D_Vp30_Angela, Pos. 89)	E Person backt Güezi (CH_Vp13_Giovanna, Pos. 27)
Possessive	Ein Designer, der seine Kleidung näht. (D_Vp28_Angela, Pos. 65)	–

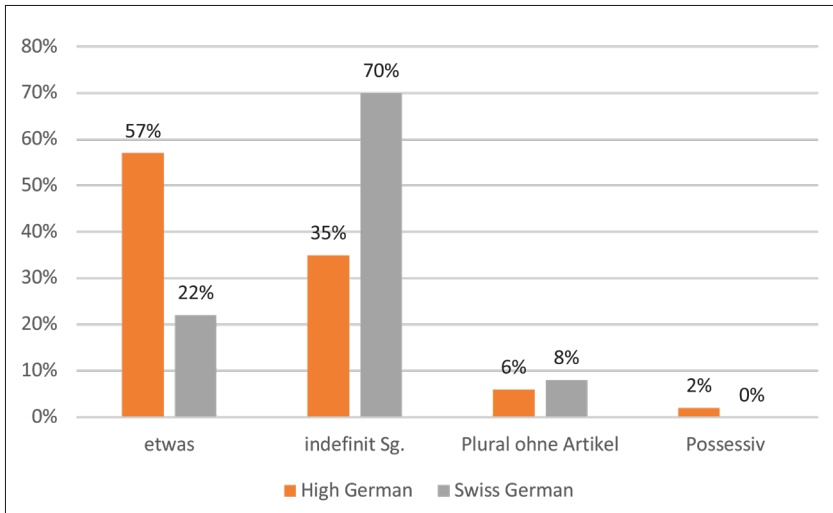


Figure 10. Syntactic embedding of the direct object in language comparison

Figure 10 shows that in the two varieties, integration via “etwas/something” and an indefinite nominal phrase are the preferred options, but other means can also be observed. In the data for Swiss High German, the expression of the direct object via “something” predominates (in 57 percent), while in Bernese the indefinite nominal phrase occurs at a rate of 70 percent.

#### 4 Summary and outlook

The results obtained in this pilot study indicate the following trends that need to be followed up: Swiss High German and Bernese differ with respect to the frequency of the *am*-periphrasis, which is higher in the Bernese dialect than in the standard variety. An individual variation emerges in both varieties, showing that the use of the *am*-periphrasis is not evenly distributed. It is the preferred form for some speakers, while others hardly ever choose it. In the case of semantic attractor effects, it was found that Swiss High German, like the Romance languages used for comparison, shows a preference to explicitly mark progression for causative events, while Bernese, like Dutch, shows no differences between activities and causative events. With respect to verbal representation, it became clear that both varieties prefer to represent events as macro-events. In Bernese, micro-events can be realized significantly more often with the *am*-periphrasis. In both varieties, the absence of an expressed object correlates with a higher use of the *am*-periphrasis which may indicate that the progressive is more often used with an activity interpretation. This



small study has illustrated that the two varieties used by the students, Swiss High German and Bernese, produce differences in form choice to some extent (more progressive in dialect), which at the same time entails differences in the representation of events (more progressive in micro-events in Bernese).

In order to obtain a more comprehensive picture of the *am*-periphrasis in German, it is planned to record data in Germany with the same stimulus material in order to explore the fine-grained differences between High German and Swiss High German on the morphosyntactic level, which have hardly been investigated so far. Recordings in Austria would also be useful to contribute to the question of the pluricentricity of German (cf. Auer, 2014) starting from progressive aspect marking. In the field of areal variation, we plan to investigate the choice of the *am*-periphrasis on the basis of the 125 survey localities of the new German language atlas SDATS (Swiss German Dialects across Time and Space) (cf. Leemann et al., in prep.). Should areal condensations in the use of the *am*-periphrasis come to light, targeted investigations would be profitable.

## Bibliography

- Auer, P. (2014). Enregistering pluricentric German. In A. Soares da Silva (Ed.), *Pluricentricity: Language Variation and Sociocognitive Dimensions* (pp. 19–48). De Gruyter.
- Behrens, B., Flecken, M., & Carroll, M. (2013). Progressive Attraction: On the Use and Grammaticalization of Progressive Aspect in Dutch, Norwegian, and German. *Journal of Germanic Linguistics*, 25(2), 95–136.
- Berretta, M. (2000) [first ed. 1993]. Morfologia. In Sobrero, A. (Ed.), *Introduzione all'italiano contemporaneo. Le strutture* (pp. 193–245). Laterza.
- Bertinetto, P. M. (1986). *Tempo, aspetto e azione nel verbo italiano. Il sistema dell'indicativo*. Accademia della Crusca.
- Bertinetto, P. M. (1990). Le perifrasi verbali italiane: saggio di analisi descrittiva e contrastava. *Quaderni Patavini di Linguistica* 8(9), 27–64.
- Bertinetto, P. M. (2000). The progressive in Romance, as compared with English. In Ö. Dahl (Ed.), *Tense and Aspect in the Languages of Europe* (pp. 559–604). De Gruyter.
- Bertinetto, P. M., Ebert, K. H., & Groot, C. (2000). The progressive in Europe. In Ö. Dahl (Ed.), *Tense and Aspect in the Languages of Europe* (pp. 517–558). De Gruyter.
- Brons-Albert, R. (1984). Die sogenannte “Rheinische Verlaufsform”: Stör mich nicht, ich bin am arbeiten! *Rechtsrheinisches Köln. Jahrbuch für Geschichte und Landeskunde* 9(19), 199–204.

- Bybee, J. L., & Dahl, Ö. (1989). The Creation of Tense and Aspect Systems in the Languages of the World. *Studies in Language*, 13(1), 51–103.
- Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The Evolution of Grammar*. Chicago University Press.
- Carroll, M., & Lambert, M. (2003). Information structure in narratives and the role of grammaticised knowledge: A study of adult French and German learners of English. In C. Dimroth, & M. Starren (Eds.), *Information structure and the dynamics of language acquisition* (pp. 267–287). John Benjamins.
- Carroll, M., Natale, S., & Starren, M. (2008). Acquisition du marquage du progressif par des apprenants germanophones de l'italien et neerlandophones du français. *Acquisition et Interaction En Langue Étrangère*, 26, 31–50.
- Carroll, M., Rossdeutscher, A., Lambert, M., & von Stutterheim, C. (2008). Subordination in narratives and macro-structural planning: A comparative point of view. In C. Fabricius-Hansen, & W. Ramm (Eds.), *Subordination* (pp. 161–184). John Benjamins.
- Carroll, M., & von Stutterheim, C. (2010). Event representation, time event relations, and clause structure: A crosslinguistic study of English and German. In J. Bohnemeyer, & E. Pederson (Eds.), *Event Representation in Language and Cognition* (pp. 68–83). Cambridge University Press.
- Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge University Press.
- Dahl, Ö. (1985). *Tense and Aspect Systems*. The Bath Press.
- Ebert, K. (1996). Progressive Aspect in German and Dutch. *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis* 1(1), 41–62.
- Ebert, K. H. (2000). Progressive markers in Germanic languages. In Ö. Dahl (Ed.), *Tense and Aspect in the Languages of Europe* (pp. 605–654). De Gruyter.
- Ferraresi, G. (2017). Ludger ist am arbeiten – Neues zur am-Verlaufsform. In Y. Ekinci, E. Montanari, & L. Selmani (Eds.), *Grammatik und Variation. Festschrift für Ludger Hoffmann zum 65. Geburtstag* (pp. 83–95). Synchron.
- Flecken, M. (2010). *Event conceptualization in language production of early bilinguals* (doctoral dissertation). LOT.
- Flick, J. (2016). Der am-Progressiv und parallele am V-en sein-Konstruktionen: Kompositionalität, Variabilität und Netzwerkbildung. *Beiträge Zur Geschichte Der Deutschen Sprache Und Literatur*, 138(2), 163–196.
- Flick, J., & Kuhmichel, K. (2013). Der am-Progressiv in Dialekt und Standardsprache. In P. Vogel (Ed.), *Sprachwandel im Neuhochdeutschen* (pp. 52–76). De Gruyter.
- Gárgyán, G. (2014). *Der am-Progressiv im heutigen Deutsch: Neue Erkenntnisse mit besonderer Hinsicht auf die Sprachgeschichte, die Aspektualität und den kontrastiven Vergleich mit dem Ungarischen*. Peter Lang.
- Gerwien, J., & von Stutterheim, C. (2022a). Describing motion events. In A. H. Jucker, & H. Hausendorf (Eds.), *Handbook of Pragmatics (HOPS) 14: Pragmatics of Space* (pp. 152–180). De Gruyter.

- Gerwien, J., & von Stutterheim, C. (2022b). Conceptual Blending Across Ontological Domains—References to Time and Space in Motion Events by Tunisian Arabic Speakers of L2 German. *Frontiers in Communication: Language Science*, 7.
- Giacalone Ramat, A. (1995). L'expression de l'aspect progressif en italien seconde langue et le rôle des propriétés sémantiques des verbes. *Acquisition et Interaction En Langue Étrangère*, 5, 47–78.
- Kempf, L. (2023). Der *am*-Progressiv in schweizerdeutschen Dialekten: Analyse einer rezenten Grammatikalisierung. In A. Speyer, & J. Dieler (Hgg.), *Syntax aus Saarbrücker Sicht 5. Beiträge der SaRDIS-Tagung zur Dialektsyntax (Zeitschrift für Dialektologie und Linguistik. Beihefte, 192)*. Wiesbaden/Stuttgart: Steiner, 71–110.
- Killie, K. (2008). From locative to durative to focalized? The English progressive and “PROG imperfective drift.” In M. Gotti, M. Dossena, & R. Dury (Eds.), *English historical linguistics 2006: Syntax and morphology 1* (pp. 69–88). John Benjamins.
- Klein, W. (1994). *Time in language*. Routledge.
- Krause, O. (1997). Progressiv-Konstruktionen im Deutschen im Vergleich mit dem Niederländischen, Englischen und Italienischen. *Sprachtypologie und Universalienforschung* 50(1), 48–82.
- Krause, O. (2002). *Progressiv im Deutschen: Eine empirische Untersuchung im Kontrast mit Niederländisch und Englisch*. Niemeyer.
- Leclercq, P. (2007). *L'influence de la L1 dans l'organisation des discours chez les apprenants avancés/quasi-bilingues: le cas de l'aspect “en déroulement” en français et en anglais* [Doctoral dissertation, Université Paris V].
- Leemann, A., Jeszenszky, P., Steiner, C., Studerus, M., & Oberholzer, L. (in preparation). *Der neue Sprachatlas der Deutschen Schweiz (nSDS)* (working title). Stämpfli.
- Lehmann, C. (2015). *Thoughts on grammaticalization*. Language Science Press.
- Levelt, W. (1989). *Speaking: From intention to articulation*. MIT Press.
- Levelt, W. (1999). Models of word production. *Trends in Cognitive Sciences*, 3(6), 223–232.
- Natale, S. (2009). *Semantische Gebrauchsdeterminanten der Verbalperiphrase stare + gerundio. Eine datenbasierte Studie*. Narr.
- Natale, S. (2018). *Informationsorganisation und makrostrukturelle Planung in mündlichen Erzählungen. Italienisch und Französisch im Vergleich unter Berücksichtigung bilingualer SprecherInnen*. Narr.
- Natale, S. (2019). Ancora sulla perifrasi progressiva. *Aspetti micro-diacronici. Vox romanica* (77), 29–42.
- Pusch, C. (2003). La grammaticalisation de l'aspectualité: les périphrases à valeur progressive en français. *Verbum*, 25: 4, *La grammaticalisation en français. Seconde partie* (pp. 495–508). Presses Universitaires de Nancy.

- Pusch, C. (2005). L'expression de la progressivité dans les français d'Amérique. In P. Brasseur, & A. Falkert (Eds.), *Langues et développement, Français d'Amérique: approches morphosyntaxiques* (pp. 159–170). L'Harmattan/Institut de la Francophonie.
- Ramelli, C. (2015a). *Die Rheinische Verlaufsform im rheinfränkischen Dialekt* [Doctoral dissertation, Universität Potsdam].
- Ramelli, C. (2015b). Grammatikalisierung im verbalen Bereich am Beispiel der Rheinischen Verlaufsform. In A. N. Lenz, & F. Patocka (Eds.), *Syntaktische Variation: Areallinguistische Perspektiven* (pp. 47–66). V&R Unipress.
- Reimann, A. (1997). *Die Verlaufsform im Deutschen. Entwickelt das Deutsche eine Aspektkorrelation?* [Doctoral dissertation, Universität Bamberg].
- Ryle, G. (1949). *The concept of mind*. Chicago University Press.
- Sasse, H. (2002). Recent activity in the theory of aspect: Accomplishment, achievement or simply non-progressive state? *Linguistic Typology*, 6(2), 199–271.
- Sichelschmidt, L., & Carbone, E. (2003). Experimentelle Methoden. In G. Rickheit, T. Herrmann, & W. Deutsch (Eds.), *Psycholinguistik: Ein internationales Handbuch*. (pp. 115–124). De Gruyter.
- Smith, C. S. (1991). *The parameter of aspect*. Kluwer.
- Strudsholm, E. (2004). Perifrasi verbali nell'italiano parlato. In Leoni, F., Cutugno, F., Pettorino, M., & Savy, R. (eds.), *Il Parlato italiano. Atti del Convegno nazionale di Napoli, 13–15 febbraio 2003* (pp. 1–10). M. D'Auria Editore.
- van Pottelberge, J. (2004). Der *am*-Progressiv: Struktur und parallele Entwicklung in den kontinentalwestgermanischen Sprachen. Narr.
- Vendler, Z. (1957). Verbs and times. *The Philosophical Review*, 66(2), 143–160.
- von Stutterheim, C., Bouhaous, A., & Carroll, M. (2017). From time to space: The impact of aspectual categories on the construal of motion events: The case of Tunisian Arabic and Modern Standard Arabic. *Linguistics*, 55(1), 207–249.
- von Stutterheim, C., & Lambert, M. (2005). Crosslinguistic analysis of temporal perspective in text production. In H. Hendricks (Ed.), *The structure of learner varieties* (pp. 1–19). De Gruyter.
- von Stutterheim, C., & Nüse, R. (2003). Processes of conceptualisation in language production: Language-specific perspectives and event construal. *Linguistics (Special Issue: Perspectives in Language Production)*, 41(5), 851–881.

---

Monique Lambert, Mary Carroll

# **Obstacles in Acquiring a Second Language: A Key to the Impact of Typological Differences on Conceptualization in Language Production**

---

**Abstract** The present chapter focuses on the means used in languages to encode concepts via grammaticalized and lexicalized forms and their role in information selection and organization in French, English and German, compared to L2 speakers of these languages. The study is based on descriptions of motion events elicited by video clips in verbal and non-verbal tasks which included retellings based on silent films by L1 speakers of French, English and German and L2 speakers in combinations of the different L1s. The findings based on these descriptions, as well as patterns of visual attention, confirm the role of crosslinguistic differences in the L1s in leading to distinctive differences in event conceptualization as well as in the distribution of attention – from the uptake of visual information to conceptual preparation. Speakers draw on language-specific schemas which guide attention to the relevant features of the stimulus in order to link the relevant linguistic categories with the associated conceptual representations. The studies show the extent to which highly competent L2 speakers draw on L1-based schemas when describing events across a large number of experimental conditions. The results not only demonstrate the degree of entrenchment of language-specific schemas but also provide insights into the factors which may impinge on the acquisition of those of the L2s. The unconscious activation of existing knowledge structures may hinder the identification of how relevant spatial concepts, their linguistic form and associated attentional processes are intertwined. Since information on a change in place with regard to the direction taken can be based on the figure in motion, (*they are heading towards y*) or features of the ground (*they are walking along x to y*),

the acquisition of the underlying event schemas in learning a second language requires adequate exposure. This will involve access to contextually grounded usage in order to gain the level of knowledge on which native-like competence is based.

**Keywords** L1 speakers; L2 learners; motion event construal; language specificity; allocation of attention; event schemata

## Introduction

This chapter presents an overview of a series of studies which were carried out in a long-term research project at the Heidelberg University on relevant factors in the second language acquisition process. Findings based on L1 crosslinguistic studies highlighted the relationship between language-specific encodings of concepts via grammaticalized and lexicalized means and the organization of information for expression. Based on these findings, focus was placed on the extent to which learners succeed in acquiring the target language-based patterns of information organization. In expressing information on events, for example, speakers map the conceptual representation of the situation onto the linguistic resources and coding patterns of the language in question. In the framework of Levelt's distinction (1989) between the level of *macroplanning* 'what to say' and *microplanning* 'how to say it', we assume that given the speed of execution required in language production, processes underlying decisions at the level of *macroplanning* as well as the level of *microplanning*, will operate in unison so as to fit the requirements of both the linguistic system as well as those of the task. The focus of interest in the study was thus placed on the interaction between cognitive processing and language specific properties (grammatical and lexical) across source and target languages on the basis of preferences in verbalization.

Starting with the early studies on language acquisition (e.g., Levelt, 1989; Slobin, 1996), the case of crosslinguistic differences with respect to the means of expression were based on the following question: When conveying information in a given context, in how far do speakers proceed on a language-specific basis? In the series of studies on the role of selective attention, for example, as elaborated under the *thinking for speaking* hypothesis (Slobin, 1996), it was assumed that the linguistic means in the language in question, both grammaticalized as well lexicalized, will direct the attention of the speaker to the associated concepts expressed in the relevant ontological domains such as *time*, *space*, and *entities*. In other words, language-specific features will be accorded salience in deciding 'what to say'.

Learners of a second language have already acquired knowledge of a system on the basis of their L1 which governs the principles underlying the use of forms, and the associated concepts, in the relevant ontological domains. This factor led to the following questions: To what extent will learners uncover the principles that drive conceptualization and information extraction in the target language? In how far will they rely on the language-specific principles of their L1 in the allocation of attention and information extraction during language production in the L2? The investigation of the task for the learner was therefore based on the crosslinguistic differences in the forms available in the languages studied when encoding information. In Talmy (2000), it was shown how languages differ in terms of the linguistic means used in encoding information on the *path* of motion, the core component of a motion event: Romance languages such as French, Spanish and Italian, are classified as *verb-framed* (V-languages) since information on the *path* of motion is expressed in the verb. This is in contrast with Germanic languages such as English, German and Dutch in which information on the *path* of motion is expressed via adjuncts and are thus referred to as *satellite-framed* (S-languages). In the *thinking for speaking* framework postulated by Slobin (1996), language specific means serve as a filtering function: S-languages, in encoding manner in the main verb, draw attention to manner of motion, while V-languages which encode path in the main verb, draw attention to path. These typological contrasts were described as a particular 'framework for schematizing experience' when *thinking for speaking*, Slobin (1996). Although early stages in language use may be language neutral or not, studies at the level of *microplanning* showed how the linguistic means available to the speaker drive the selection and organization of information.

The crosslinguistic research surveyed below covers investigations on the implications of both grammaticalized as well as lexicalized contrasts at different levels in the production process. The focus of analysis concerns the role of linguistic forms, with their associate concepts, in driving attention to aspects of a situation that are readily encodable when verbalizing an event. Within the large body of research carried out by the Heidelberg group, the present overview includes L1 speakers of French, English and German as well as highly competent L2 learners with these specific L1s and L2s. The objective was to pinpoint the obstacles learners may still face at advanced stages of acquisition with regard to the question of full competence. This requires insight into the nature of the knowledge activated in language processing at the conceptual level. Regarding language acquisition, the findings provide a basis for discussion of the factor of frequency, which is viewed as a driving force in the usage-based approach along with the relevance of 'conceptual transfer', as postulated by Jarvis & Pavlenko (2008).

The following sections are structured as follows: Part 1 describes the underlying theoretical framework and related studies. This is followed in Part 2 with an overview of the studies on motion event construal that combine linguistic analysis and psycholinguistic procedures. In Part 3, a range of specific studies, which were based on the description of single events elicited by video clips, are compared with events verbalized in a narrative context, as in retellings of silent films.

## 1 Theoretical Underpinnings

### 1.1 Language production models

The series of studies outlined below draw on models of language production with an initial level of ‘conceptualization’, in which content is selected for expression, followed by its ‘formulation’ in linguistic form and the final process of ‘articulation’ (cf. Levelt, 1989). The initial stage at the conceptual level covers the stage of ‘macroplanning’ as related to processes involved in information selection and organization. This is followed by ‘microplanning’ where the conceptual content selected for expression is brought into perspective and assigned a particular information structure on the basis of language specific requirements. In this framework, the initial stage of macroplanning is viewed as language neutral, while crosslinguistic differences are confined to the microplanning stage only.

On the whole, studies on the encoding processes involving manner and path show that speakers of different languages focus on different components of motion events when *thinking for speaking*. This is viewed as based on the activation of language-specific patterns of attention.

### 1.2 L1 and L2 Acquisition

Empirical research on L1 and L2 acquisition provides insights into the potential interaction between planning processes at macro and micro levels in language production. Slobin (1996) states that in the course of extensive exposure and use of the L1, children develop preferences in the direction of attention to the relevant conceptual domains and associated information when representing situations or events. These ‘habits of thought’ are acquired in conjunction with the acquisition of the specific formal means over a long-term process of cognitive and linguistic development (Hickmann et al., 1998; von Stutterheim et al., 2012 and overview in Özçalışkan & Emerson, 2016). L2 learners have the ‘habits of thought’ of their source language, in Slobin’s terminology, and the question arises as to how these intervene when



processing input from the target language. The well-established impression shared by native speakers that something may be unconventional in the verbal production of even highly fluent L2 speakers may be attributed to the activation of L1-based attentional patterns when *thinking for speaking*.

With regard to the question of the L2 speaker's abilities, and the conditions which may, or may not, lead to the acquisition of the relevant range of knowledge 'when *thinking for speaking*' in the target language, we now turn to usage-based theories on L2 acquisition (Ellis, 2008; overview in Kartal & Sarigul, 2017). In this context, the extent to which speakers are exposed in everyday language to linguistic expressions and their combinations is rated as a major driving force. Second language learning is viewed as based on implicit cognitive mechanisms such as sensitivity to frequency of occurrence, along with strings of occurrences in the L2, which allow the extraction of statistical regularities and inductive generalizations of collocational dependencies. According to Wulff & Ellis (2018) on the role of the L1, L2 acquisition is viewed as filtered through the lens of the L1 with beneficial effects when features are shared by both the source and target language and negative effects when features differ. In this context two questions arise: Could the level of 'non-nativeness' in advanced learners' productions arise from difficulties in extracting statistical regularities based on *preferred* usage of linguistic forms in context, in contrast to formal regularities? Could frequency of exposure have a limited impact on deeply entrenched and automatically activated L1 'habits of thought' when *thinking for speaking* in L2? These questions would require the investigation of how selective attention operates on cognition both on a linguistic as well as on a non-linguistic level. The crosslinguistic studies on motion events presented here provide insights that are based on language effects.

### 1.3 Studies on Motion Events

In the theoretical framework initially proposed by Talmy (1975, 2000), motion events involve the following conceptual/semantic categories: *figure*, *motion*, *ground*, and *path* whereby *manner of motion* and *cause* constitute co-events. Given the fact that the *path* traced through space by a *figure* in motion is viewed as the central conceptual component of a motion event, languages can encode information which comprises features of the *ground* traversed by the figure (source, trajectory, goal) as well as the *orientation* of the figure. Languages vary in the means which encode information on the *path* of a motion event: In Romance languages, information on the path of motion is expressed in the verb and are thus termed *verb-framed* ('a man *enters* the building') in contrast to *satellite-framed* in Germanic languages ('a man is walking *into* the building').

Depending on whether information on the path of motion is encoded in verbs versus satellites, different components of the situation will be expressed explicitly, or remain implicit. Crosslinguistic studies on the scale of such differences showed variation, both across typologically related languages as well as within the language groups. In Romance languages, where *path* is typically encoded in the verb, studies showed that speakers also use manner verbs at relatively high frequencies (Stringer, 2005; Cardini, 2008, 2012; Kopecka, 2009; Slobin et al., 2011). This led to distinctions based on the actual salience of the concepts *path* or *manner* across languages (Hickmann & Robert, 2006; Slobin 2004, 2006). Although mainly satellite-framed, English also shows use of path verbs (Carroll et al., 2012). Other studies included additional criteria covering *manner of motion* (Slobin, 2006; Pourcel, 2005; Feist, 2016) with the type of 'figure' in motion *animate/inanimate entity* (Pourcel & Kopecka, 2005). A further level of analysis in the study of events focused on the role of *bound- edness* (von Stutterheim et al., 2012; Flecken et al., 2014) as well as *aspectual* categories (events marked as *ongoing*), when expressing information on a change in place (von Stutterheim et al., 2017). The relevant typological differences were compared across a range of typologically diverse languages (Levinson, 1996, 2003; Romance versus Germanic languages: Athanasopoulos et al., 2015; Hickmann et al., 2008; Durst-Anderson et al., 2013; English versus Greek: Papafragou et al., 2010).

The initial focus on the components 'path and manner' was followed by an extension of the analyses to a comprehensive view on events at the overall level of event construal, (von Stutterheim et al., 2012; Bylund & Athanasopoulos, 2013; von Stutterheim et al., 2017; von Stutterheim et al., 2020). A further line of research involved the implications of the different patterns of lexicalization on processes underlying motion cognition which could provide insights on the level at which language specificity plays a role. This was carried out on the basis of verbal tasks which were coupled with non-verbal experiments (triad categorization tasks, similarity judgments, memory tests, verbal interference, eye tracking). On the whole, results concerning verbal description tasks were consistent with previous findings, but non-verbal tasks produced conflicting findings that could be related to task choices, stimuli, instructions etc., (overview in Pavlenko, 2014).

Studies on L1-specific effects at the conceptual level in a second language take at its starting point cognitive differences in monolingual populations in a specific domain. They examine the extent to which language-specific cognition in the same domain shows (partial) L1 influence and (partial) resistance to L2 influence. The findings also are not consistent, however. Studies on learners at early stages conducted by Cadierno (2004), for example, showed that Danish learners of Spanish manifest L1 influence in their description of motion events providing elaborate path information in contrast to L1 Spanish

speakers. L1 effects were also identified in boundary crossing contexts where learners used manner verbs (Cadierno & Ruiz, 2006). A study on the impact of the level of proficiency at early stages (Athanasopoulos et al., 2015) tested English learners of German with different lengths of exposure to the target language. The study was based on the finding according to which speakers of aspect languages are less likely to attend to endpoints (goal) compared to speakers of non-aspect languages. The categorization tasks used motion events with different degrees of goal orientation. Results show L1 behaviour patterns for learners with medium exposure and a shift to L2 behaviour patterns with the highly exposed learners. Other studies on advanced L2 learners, however, showed persistent L1 effects on motion event construal (Flecken et al., 2015; Bylund & Jarvis, 2011; Bylund & Athanasopoulos, 2015; Treffers-Daller & Tidball, 2015; Pavlenko & Volynsky, 2015; von Stutterheim et al., 2013). The same holds in experiments in which verbalizations were coupled with speech onset times (SOTs) and eye tracking procedures. In sum, the crosslinguistic studies show how relevant language-specific contrasts tend to play a significant role at the preverbal phase in language production. The next section provides further descriptions of some of the studies mentioned above which were carried out by the Heidelberg group. Focus is placed on cognitive differences across L1 speakers in examining the extent to which the representation of motion events by advanced users of an L2 is resistant to L2 influence and remains influenced by L1 language-specific processes.

## 2 Processes of conceptualization in L2 production

The empirical studies address the level at which language-specific structural properties influence processes in L2 construal of motion events: which processes (L1 or L2 based) operate at the encoding stage of microplanning as well as at the macro level, and to what extent, as reflected in attention allocation, event segmentation and information selection.

### 2.1 The Impact of Aspectual Categories on the Construal of Motion Events

A series of studies tested the role of ongoing aspect on the selection and organization of event information in grammaticized aspect languages versus non-aspect languages. The relevance of this variable is based on early studies on narrative texts produced by L1 speakers of French and German and French and German learners of English (Carroll & Lambert, 2006). It was shown that while speakers of languages that do not encode aspectual distinctions on a broad scale (German and French) tend to represent events which occur in

sequence as bounded with a possible endpoint, speakers of English, given the presence of aspectual means, may view an event as bounded, or ongoing. In the case of event cognition, speakers of non-aspect languages may be more likely to adopt a holistic perspective in motion events viewed as bounded. In contrast, speakers of aspect languages may, in relevant contexts, decide whether an event is ongoing or completed. Consequently, they may segment an event into phases as it unfolds and attend to the current ongoing phase.

This led to testing this assumption in a more controlled research design. Speakers of different languages (+/- aspectual distinctions) were asked to describe video clips covering a range of everyday motion situations which were distinguished by the presence of an endpoint. The clips included scenes in which the endpoint of the trajectory is reached (control item), while a further set (critical items) featured scenes where the endpoint had to be inferred, and differed in the degree of goal orientation from an intermediate degree (a man driving a car along a road toward a village) to a low degree (two women walking in a park with a bench in the distance). The range of scenes also included situations with distractors showing activities with no inferable endpoint. Participants (20 per group) were asked to answer the question *What is happening?* and their responses were recorded. In addition to analyses of the linguistic data, additional methods (e.g. speech onset times SOTs) were implemented which could provide insight into possible differences during the conceptualization phase in speech production. It was assumed that longer delays in speech onset may reflect the tendency to include an endpoint in languages that do not encode aspectual distinctions (event is ongoing). Eye tracking measurements provided relevant information at the preverbal phase while the event unfolds: fixations patterns were in line with the tendency to encode an endpoint. Variations at this level thus reflect language-specific effects in language production, as shown in a subsequent set of studies. The experimental studies included verbal as well as non-verbal processing (e.g., von Stutterheim & Nüse, 2003; von Stutterheim et al., 2012; Flecken et al., 2014; Flecken et al., 2015); Gerwien & von Stutterheim (2018). For example, speakers of non-aspect languages (German, Norwegian, Swedish, French) tend to represent events under a holistic perspective by mentioning the inferred endpoint. This is not the case with speakers of aspect languages (English, Russian, Arabic and Spanish) in which events are verbalized as 'ongoing'. The frequency with which endpoints are mentioned is significantly lower compared to speakers of non-aspect languages. Findings based on SOT measurements also pinpoint systematic contrasts in time span for participants of non-aspect languages while waiting for a potential goal / endpoint before speaking. In contrast, speakers of aspect languages start speaking as the situation unfolds (von Stutterheim & Carroll, 2006). These differences reflect language-specific effects on processes prior to verbalization.

Given these findings a study carried out by von Stutterheim & Carroll (2006), investigated the impact of aspect on the representation of events by L1 German-L2 English and L1 English-L2 German learners<sup>1</sup>. Options based on comparisons with learners and speakers of the target language for the same tasks reflect significant L1 influence in the verbalizations. Although L1 German-L2 English speakers use the progressive form, they still include endpoints when describing the critical items. This is not typically the case for speakers of English when events are presented as ongoing. An endpoint may or may not be included, depending on its relevance in the given context. Patterns in visual attention, measured by SOTs, may correspond to those which are typical of the learners' source language. This was the case with L1 English-L2 German learners. The patterns for L1 German-L2 English were closer to L1 English, however. This difference may be attributed to the status of the progressive in English ('event is ongoing'), given its frequency. In German, there is no form which explicitly express the concept 'holistic' for events (von Stutterheim, 2003; von Stutterheim et al., 2009). An overt form (the progressive) may make distinctions in event conceptualization more salient for the learner and contribute in identifying the concept and its function.

## 2.2 Conceptual Domains Underlying Key Concepts in Describing the Trajectory

The study conducted by Carroll et al. (2012) investigated the type of concepts encoded in verbs, prepositions, verb particles and adverbs in motion event descriptions of the trajectory with regard to the basic conceptual domains from which they are derived in English, French and German. The relevant contrasts are as follows: in French the concepts which are activated when describing the event are derived from features which relate to the direction taken by the 'figure in motion' as well as the ground, albeit to a lesser extent. This is in contrast to German in which 'contours of the ground' are predominant. The study, which included two groups of advanced learners of English, with French and German as their source languages, compared the acquisition of the relevant concepts when tracing the trajectory through space. The study was based on video clips featuring motion events. One set of scenes showed an entity in motion on the section of the trajectory close to an evident goal (+ Goal) in contrast to a set which focused on the stretch leading to it (+ Ground). The goal was less evident in the latter case. The contrast would indicate the extent to which the spatial concepts used by the learners

---

1 The participants (20 per group) were all highly competent users of their L2s assessed by tests and length of institutional language studies and eventually studies abroad.

would focus on contours of the ground, since attention to an endpoint is also possible without focusing on information at this level.<sup>2</sup>

The findings, based on statistically validated preferences, provide clear evidence of the relevant language-related contrasts. In the clips showing an evident goal, spatial components are encoded in L1 French on the basis of verbs as follows:

- (i) verbs relating to the figure in motion, its trajectory and goal *se diriger vers x* ('to direct oneself toward x'), *s'approcher de x* ('to approach x'),
- (ii) verbs that relate to the orientation of the figure in motion *tourner* (to turn).

As indicated above, the relevant contrast in language acquisition is that the concepts encoded in the verb relating to *the figure* in motion are used frequently in French, in contrast to features of *the ground*. In English, by contrast, the dominant forms which conjoin with manner of motion in the verb (e.g., *to run*; *to walk*) relate to features of the ground (*through*, *along*, *around*, *over x*). In the clips showing an extended trajectory and no evident goal, manner of motion is expressed in the verb in French. These verbs typically conjoin with adjuncts which express the location of the figure in motion (*sur la route*, 'on the road'), but not direction. There are few forms in French that relate to specific contours of the ground, such as *le long de*, for example, and overall use is low. This is the main contrast to English. Manner verbs predominate (e.g., *to walk*, *to drive*) and typically combine with spatial information based on features of the ground (*they are driving along a road around a village*).

The comparison of L1 speakers of French and English showed how the crosslinguistic variation in the means available to the speaker lead to differences with regard to the features of the scene which come under focus in the stimuli. Depending on the prominence of the goal, the following patterns in event construal are activated in French: (i) with a prominent goal and a short trajectory, speakers encode information on both the direction taken with verbs relating to the figure in motion and the goal; (ii) in the absence of a salient goal, the dominant pattern is to refer to manner of motion with mention of the location of the entity in motion (*marcher sur la route*), but not its direction. In English, manner of motion predominates in the information encoded in the verb. This conjoins with spatial concepts that express the path

---

2 The range of languages in the overall project include Romance languages (French, Italian Spanish), Germanic (English, German, Dutch, Norwegian) and Semitic (Modern Standard Arabic) and learner languages with L1s and L2s in the above cited languages.

taken and are based on contours of the ground (e.g., *along the street, around the corner, over the bridge*).

Learners of English with L1 French, for example, acquire use of manner verbs as well as the deictic verbs. But in contrast to L1 speakers of English, the learners do not use the wide range of spatial concepts (e.g., *along, around*) which are derived from contours of the ground and shape the trajectory leading to a goal. When no goal can be detected in the stimuli, the scene is represented by the learners as ‘an entity moving in a certain manner located at a place (*une voiture roule sur la route; a car drives on the road*), as in L1 French. They maintain a focus of attention based on their source language along with the relevant spatial concepts and their linguistic form. Depending on the conditions under which the language can be learned (classroom; lack of immersion with native speakers), the knowledge whereby core grammaticized concepts drive the focus of attention in language processing may prove difficult to acquire.

Based on the findings presented above, a follow up study was carried out (Flecken et al., 2015) investigating the level at which language specificity operates on L1 patterns and the possible consequences for L2 learners. The study included L1 French-L2 German learners as well as L1 French and L1 German speakers. Their descriptions of the events were coupled with both eye tracking recordings as well as measurements of speech onset times (SOTs). The analysis was based on the following hypothesis: Given the role of the L1, speakers of French will focus on the direction taken by the entity in motion in relation to a (potential) goal (e.g., *to head toward x*) when this is evident, and on its location when it is not. Speakers of German will relate to features of the ground through which the entity is moving (e.g., *über [over], unter [under] entlang [along]*).

The findings show that based on concepts which indicate the direction taken, as expressed by verbs in L1 French, speakers of French allocate a higher level of attention to the entity in motion, as well as to the possible endpoints, before speech onset. This is not the case for L1 speakers of German who show higher levels of focus on features of ground, compared to the French speakers. The use of forms that indicate direction (e.g., *über [over], unter [under], entlang [along]*) via features of the route is high. Although the use of manner verbs by the L2 speakers of German with L1 French were similar to the speakers of L1 German, this is not the case for the patterns of attention to features of the ground by the L2s, prior to speech onset. These were similar to those of native speakers of their L1 (French), as shown by the eye tracking recordings. The L2 speakers allocate attention to the moving entity to a higher degree, in contrast to the L1 speakers of German. The L2 pattern is thus in line with L1 French.

The findings for this group of L2 speakers show how the process of conceptualisation when ‘thinking for speaking’ may be mediated by the source

language, as illustrated by the specific components of the scenes which are placed under focus when preparing to speak (entity in motion / specific features of the ground). They confirm the role of attentional processes and associated patterns of conceptualization when processing information for expression. This may affect acquisition of the knowledge underlying event representation in the target language when learning an L2.

### 2.3 Formation of Event Units

Although the empirical study by Gerwien & von Stutterheim (2018) did not include L2 users, it is relevant to L2 acquisition given the investigation of language effects at the earlier stages of event segmentation and provides insights on the degree of entrenchment of L1 patterns of conceptualization. The study compared the extent to which the relevant structural differences between French and German influence the way in which the continuous flow of information is segmented into event units. In French, where motion from one place to another by the entity in motion is predominantly expressed by path verbs, segmentation will correlate with changes in orientation *descendre*, *tourner*. In German, by contrast, motion from one place to another is typically expressed in terms of manner of motion. Information on the path taken is encoded in forms which may cover a number of path segments (*sie laufen eine Straße entlang und über eine Brücke auf dem Weg nach Hause*; they walk 'along' a street and 'over' a bridge 'on the way' home). Given these basic contrasts that relate (i) to the role of figure in motion and (ii) the ground traversed, speakers of French and German may draw on representations of a motion event at a conceptual level (event schema) which differ in the focus placed on features of the event. The experiments were based on video clips showing a set of situations in which an entity moves along a path with changes in direction (up, down), orientation (left, right) followed by motion to a goal. These were accompanied by control clips which varied (i) in the degree of prominence of these changes in direction along with (ii) clips with no change in direction. In the first experiment, participants were asked to describe the scenes shown in the videos. This was followed by the second experiment, based on the Newton task (1973), with a second group of participants who were asked to press a button whenever they noticed a change in situation.

In the verbal task, speakers of French tend to describe a motion event with more than one sentence, in accordance with change in direction (new verb). This is in contrast to German speakers who typically describe a motion event based on one sentence only with the relevant nominal phrases. The results for the non-verbal segmentation task are in accordance with those of the verbal task: The speakers of French were more likely to mark an event



boundary at the point where a change in orientation/direction occurred in the stimulus. This pattern was not observed since changes in orientation is expressed by prepositions which can be included in one sentence for the speakers of German.

The findings for both languages show how the activation of processes in both verbal and non-verbal tasks are guided by the conceptual categories that are obligatory and frequently accessed by speakers of a specific language. The impact of language-specific structural properties in the process of segmentation reflects the degree of entrenchment of L1-based effects. This pinpoints the difficulties with which learners have to deal in acquiring the processes in operation at this level.

### 3 Motion Events in a Narrative Context

The following research questions whether the different language specific patterns described above are robust and activated in different types of communicative situations when learning and using a second language. This led, in a first step, to investigate motion events produced in a narrative task (the retellings of a cartoon, Reksio) and see whether L2 learners attend to aspects of motion events that are specific in their target language or remain influenced by their L1 preferences.<sup>3</sup> In order to diversify the type of stimuli, the data further included retellings of an animation film, Quest, as well as episodes of the film Modern Times featuring real persons. The study is based on L1 speakers of English and French as well as highly competent L1 French-L2 English learners (N=20 per group). To test the implication of tasks conditions, motion event descriptions produced in the semi-controlled task (the retellings) were compared to those produced in the description of a selection of video clips used in the Carroll et al. study (see 2.2) which best matched the scenes selected in the films.

The requirements of a narrative task are complex at a superordinate level: Based on mental representations of the perceived unfolding film, motion events are memorized, retrieved, selected and further organized so as to satisfy constraints based on coherence and fit the narrative format. As reflected in the number of occurrences of motion events per scene (see Tables), these constraints lead to individual variation in the information selected for expression in the retellings, both in the L1 and L2 data.

The unpublished study presented here comprises retellings of the cartoon (Reksio, in which a boy with a dog go skating on a lake where the water is only partly frozen and the boy falls in), an animation film (Quest) which

---

3 Presented at Eurosla 22 conference (2012)

features a sandman searching for water in a series of hostile settings, and an episode from the film ‘Modern Times’ (Chaplin), in this case with two main protagonists. The films were first shown in their entirety and repeated once more episode by episode in order to support the speaker when re-telling the events. Participants were recorded one by one and the accompanying experimenter interfered only when it was necessary to show encouragement. The retellings were elicited by the question *What happened?* for both the L1 and L2 groups and this elicitation question was repeated for each episode. In order to compare both the retelling tasks and the video description task, the analyses focused only on goal-oriented motion events which differed according to the length of the trajectory to a potential goal. The first category comprised events with a short trajectory and an evident goal *il se dirige vers la source de l’eau* (he heads towards the pool of water) with a second category showing events with a long trajectory and no evident goal (in French: *ensuite une fille marche dans la rue* [a girl walks in the street]) (in English: *walks along the street*). The data comprised ten events (five in each category). The descriptions of video clips which were added to this group included scenes with a prominent goal and a short trajectory (4) as well as scenes with a distant potential goal (5). Examples from scenes with a prominent goal from the database of the retellings are as follows: *a person walking towards a large pool of water; a man walking to a shipyard* and from the descriptions of video clips they include *a woman walking along a stretch of road towards a car; someone walking across the courtyard or square to a phone box*. Examples illustrating scenes with a distant potential goal are *a woman walking down a street; a man walking along a road* from the retellings and *two nuns walking down a lane* and *a person is walking around a corner of a street* from the videoclip descriptions. In line with the coding criteria used in the study by Carroll et al. (2012), information encoded in the verb covers the following categories: manner of motion, direction, the deictic ‘come’ and ‘go’. Information coded in adjuncts relate to the goal (*towards, to*); the path (*along, over, down*) and in some cases location (*in*). The category ‘no adjunct’ was also included. Although the initial overall instruction was in the past tense, the retellings were grounded in the present across all the languages. This was also the case for the descriptions of video clips by L1s and L2s from which we draw our selection of scenes. English speakers must decide, when mentioning an event, whether it requires an ongoing perspective (*they are running to the train*), or not (*they run to the train*). Scenes without an evident goal, for example, were generally encoded as ongoing (*-ing*) in L1 English but less so in the scenes where the trajectory was short. L2 learners used the present and the progressive in both scenes but the options did not show any consistency which could be linked to the type of motion event. The numbers in the following tables are based on the sum of motion events expressed in the three retellings for the event types

(+ Goal) and (– Goal). In a first step (Table 1) the frequency of verbs encoding Manner vs Direction vs Deictic ‘go’ in the retelling tasks was compared to those based on the video clips. Adjuncts in (Table 2) are coded as follows: no adjunct, adjuncts referring to overall location (Loc): *he is walking in the park il marche dans la forêt*, to contours of the ground *he is walking along the road, il marche le long de la route*, to goal direction (To): *they are walking to(wards) a house ils vont vers la maison*, to goal location (At): *il va à la maison*. In a second step, frequencies of the spatial forms used in the retellings by the L2s were compared with the target and source language speakers for the same variables: Verbs (Table 3) and Adjuncts (Table 4).<sup>4</sup>

### 3.1 Comparisons of the L1s across the Tasks

Tables 1 and 2 present the relative frequencies of verb types and adjunct type used by the L1s in the film retellings and in the description of the video clips. Frequencies are based on the total number of occurrences (first column) produced in the retellings per group of narrators (15 participants) in the overall (+ Goal) and (– Goal) scenes.

Table 1. (+) Goal-oriented motion events: Verb form in L1 French and L1 English (retellings and clips)

Retellings	Manner	Direction	Deictic
L1 Fr n=45	2.2%	60.9%	32.6%
L1 Eng n=46	34.8%	2.1%	63.1%
<hr/>			
Clips	Manner	Direction	Deictic
L1 Fr n=80	3.7%	58.1%	38.1%
L1 Eng n=91	87.7%	2.9%	9.5%

Table 2. (+) Goal-oriented motion events: Adjuncts in L1 French and L1 English (retellings and clips)

Retellings	No adjunct	LOC	Contours	TO Obj	At Obj
L1 Fr n=45	0	0	2.0%	42.8%	55.2%
L1 Eng n=67	0	0	43.3%	56.7%	0

4 See L1-L2 comparisons for video clip descriptions in Carroll et al. (2012).

Table 2. (continued)

Retellings	No adjunct	LOC	Contours	TO Obj	At Obj
Clips	No adjunct	LOC	Contours	TO Obj	At Obj
L1 Fr n=80	1.2%	27.5%	2.5%	61.3%	7.5%
L1 Eng n=91	3.3%	6.6%	36.3%	64.4%	2.2%

There were no differences in the forms used across the data sets for L1 French (see Carroll et al., 2012): Verbs expressing direction, along with the deictic *va*, predominate, as expected, whereby use of the different prepositions is linked to the type of verb: *se diriger* (to head for) combines with *vers* (toward) while *aller* (go) can combine with forms expressing either direction *vers* (toward) or location *là où*.

Table 3. (-) Goal-oriented motion events: Verb encoding in L1 French and L1 English (retellings and clips)

Retellings	Manner	Direction	VA/GO
L1 Fr n=45	78.4%	17.6%	3.9%
L1 Eng n=46	96.1%	3.9%	0
Clips	Manner	Direction	VA/GO
L1 Fr n=80	85.2%	0	14.8%
L1 Eng n=91	96.3%	0	3.7%

Table 4. (-) Goal-oriented motion events: Adjuncts in L1 French and L1 English (retellings and clips)

Retellings	No adjunct	LOC Global	Contours	TO Obj
L1 Fr n= 62	19.4%	76.5%	1.6%	3.2%
L1 Eng n=55	14.5%	12.7%	69.1%	1.8%
Clips	No adjunct	LOC Global	Contours	TO Obj
L1 Fr n=64	15.6%	71.9%	7.8%	4.7%
L1 Eng n=70	2.5%	20.2%	67.1%	6.3%

In English, the verbs expressing manner of motion, as well as the deictic *come* and *go*, combine with forms relating to contours of the ground (*along, over, on, up, down*) as well as direction to a goal: *the girl is walking along the street towards the bakery*. In contrast to English, when goals are visible, verbs in French relate to the direction taken based on the orientation of the entity with respect to a goal point. When a goal is hardly visible, manner verbs are used consistently with mention of the location of the entity only. We can conclude that in both tasks, speakers activate conceptual representations of the events which guide the allocation of attention to the relevant conceptual components of the motion event. Significantly, this provides a clear example of the type of knowledge which learners of a second language have to acquire.

### 3.2 Comparison between L1 and L2 Speakers

The following tables based on the retelling tasks provide an overview of the relative frequencies for the types of verbs (Table 5), and adjuncts (Table 6) used in English and French, along with the L1 French-2 English learners.

Table 5. (+) Goal-oriented motion events: Verb encoding in comparison between L1 and L2 speakers (retellings)

Retellings	Manner	Direction	VA/GO
L1 Fr n=45	2.2 %	60.9 %	32.6 %
L1 Eng n=46	34.8 %	2.1 %	63.1 %
L1 Fr-L2 Eng n= 46	21.1 %	31.6 %	47.4 %

Table 6. (+) Goal-oriented motion events: Adjuncts in comparison between L1 and L2 speakers (retellings)

Retellings	No adjunct	LOC	Contours	To Obj	At Obj
L1 Fr n=45	0	0	2.0 %	42.8 %	55.2 %
L1 Eng n=67	0	0	43.3 %	56.7 %	0
L1 Fr-L2 Eng n=48	4.2 %	4.2 %	4.2 %	81.1 %	6.2 %

The use by L2 learners of English of the deictic verb *to go* (47.4%) may be significant in that it not only has a directional component but is also highly frequent in English (63.1%). Directional verbs (*to head for x, to reach x*) are low in frequency for L1 English (2.1 %) but occurrences for the learners add

up to 31.6%. In contrast to English, where information on the trajectory is encoded in forms which are based on contours of the ground (*along, around, over, under*), the rate of occurrence of these forms in the learner group is markedly low. Table 7 and table 8 cover the relative frequencies for the set of stimuli which feature a prominent trajectory but no evident goal.

Table 7. (-) Goal-oriented motion events: Verb encoding in comparison between L1 and L2 speakers (retellings)

Retellings	Manner	Direction	VA/GO
L1 Fr n=51	78.4%	17.6%	3.9%
L1 Eng n=51	96.1%	3.9%	0
L1 Fr-L2 Eng n=51	80.4%	7.8%	3.9%

Table 8. (-) Goal-oriented motion events: Adjuncts in comparison between L1 and L2 speakers (retellings)

Retellings	No adjunct	LOC Global	Contours	TO Obj
L1 Fr n=62	19.4%	76.5%	1.6%	3.2%
L1 Eng n=55	14.5%	12.7%	69.1%	1.8%
L1 Fr-L2 Eng n=42	38.1%	42.9%	11.9%	2.3%

In the scenes with no evident goal, the use of manner verbs is prominent across all three groups. In the L2 data however, these forms do not conjoin with ground-based concepts, even though contours of the ground are visually prominent. The occurrences of adjuncts, which in L1 English express direction (e.g., *along, around, over*), is low at 11.9% for the L2s compared to a rate of occurrence of 69.1% for L1 English.

We can conclude from these comparisons that the experimental conditions with their variance in cognitive load do not significantly alter the options shown in the descriptions of the motion events. Whether elicited online by a visual stimulus or based on a mental representation stored in memory and reactivated to fit in a narrative format, speakers of English and French show differences in their options which reflect the impact of the structural properties of the respective L1 languages. They remain globally consistent in their options across tasks. In both tasks (see findings in 2.2 above) the event descriptions by the L2 speakers reflect the impact of the structural properties of their L1 which are independent of task conditions. These converging results manifest the degree of entrenchment of the underlying knowledge based on the L1.

## 4 Discussion

Crosslinguistic differences in the patterns of conceptualization observed in the L1 studies on motion events, based on verbal and non-verbal tasks, reflect the impact of language-specific structural properties at the pre-verbal stage of language production. The factors in question relate to visual uptake, event segmentation as well as information selection. As assumed in Lambert et al. (2022), in order to carry out the verbal task of describing a motion event speakers draw on language-specific event schemas. They function as a guide in the allocation of attention to the relevant features of the stimulus when linking linguistic categories with their associated conceptual representations. Speakers of a given language community develop this schematic knowledge by repeated exposure to the way in which attention is focalized on specific conceptual features of events, and the associated ontological domains (*time*, *space*, and *entities*) when using linguistic structures in a verbal task. When describing a change in place, speakers of French focus on spatial concepts which are based on the figure in motion. Speakers of English, in contrast, focus on concepts based on features of the ground. These event schemas relate to the relevant cognitive processes as follows: They serve in integrating the associated constraints at the level of contextual information with regard to the structural means which are adequate for a given task. When activated they provide automatic guidance in the selection of information, along with the means of expression required, with the appropriate speed of execution in speech production. This knowledge is shared by native speakers and forms part of their cognitive-pragmatic knowledge (Gerwien & von Stutterheim, 2022). The obstacles which late learners of an L2 encounter can be located within this framework.

The task for the L2 learners is not confined to the acquisition of the relevant forms with their syntactic properties. They must also learn how a given situation/scene is typically represented at a conceptual level and expressed verbally in the target language. Although the L2 participants in the studies presented here no longer have difficulties with respect to the formal means of expression, the findings, both verbal and non-verbal, pinpoint the level at which obstacles persist. Divergences from the target language speakers can be viewed as located at the pre-verbal stages when preparing a message for expression: Convergences with speakers of their source language at this level point to the activation of processes that guide both the focus of attention as well as information selection. The processes in question are rooted in the pragmatic use of the available means within a language community.

Why should this be the case? As discussed in von Stutterheim & al. (2021), we consider the role of frequency of forms as a key factor in acquisition. L2 speakers have been potentially exposed to highly frequent concepts (*along*, *around*,

over) which are derived from features of the ground (e.g., *she is walking down the street; he is walking across the square to the shipyard*). Yet, occurrences of these forms are markedly low in the present L2 data base. In the contexts where ground-based information is typically verbalized in shaping the trajectory in English, the learners combine manner verbs with a specification of the ‘location’ of the entity, with no mention of its direction, as in French. This shows how frequency of exposure is far from automatic in driving acquisition since combinations that are typical in many contexts also remain unnoticed and resist acquisition. Very low frequencies do not seem to play a role either: In English the verbal means which refer to the direction taken (e.g. *to head for, to approach, to reach*), although relatively infrequent in use, are nevertheless used by the L2 speakers to a significantly higher degree compared to the L1 English speakers (respectively 02,1 vs. 31,6).

We can conclude from these findings that the expressions used in describing motion events, with their underlying concepts which are based on the L1 (‘figure in motion’ versus ‘features of the ground’), take precedence over the frequency of occurrence of forms in the target language. The statement in Wulff & Ellis (2018) confirms the non-automaticity of frequency effects: “Since everything is filtered through the lens of the L1, not all of the relevant input is in fact taken advantage of”. This statement holds at a level of event construal when formal means no longer pose a problem for the L2 groups observed.

An L1 effect in the allocation of attention was also observed at early stages of conceptualization in the eye tracking data. The L2s prioritize the moving entity in line with the pattern obtained for L1 French. This is in clear contrast to L1 English speakers who prioritize the domain of space.

In summary, given the level of representation at which language-specific schemas in event construal are stored, the present findings indicate the nature of the deep-seated ‘obstacles’ which the learners face. Given the level of abstraction of the underlying event schemas, as well as the time span over which they develop, the long-standing notion of ‘transfer’ (Jarvis, 2007) may be inadequate. In order to develop the unconscious attentional processes required, the L2 speakers must derive the relevant schemas from inferences and generalizations which are based on preferred uses in context. Adequate access to this type of knowledge would improve the basis for the development of the language-specific patterns of conceptualization to a large degree at least.

In conclusion, the findings provide noteworthy insights into the implication of language in higher order processes and how processes interact at both the macro and micro level. Concerning L2 learners, it remains clear in the findings documented above that the crosslinguistic divergences can be traced to the continuing impact of structural properties of the L1 since



their verbal responses reflect clear similarities with those of their L1. Overall, the studies documented here provide insight into the nature of the deep-seated obstacles that learners face before they can automatically attend to, and represent, events based on the relevant structures and processes in the target language.

## References

- Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., Riches, N., & Thierry, G. (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science, 26*(4), 518–526.
- Bylund, E., & Athanasopoulos, P. (2015). Televised Whorf: Cognitive restructuring in advanced foreign language learners as a function of audio-visual media exposure. *Modern Language Journal, 99, Supplement 1*, 123–137.
- Bylund, E., Athanasopoulos, P., & Oostendorp, M. (2013). Motion event cognition and grammatical aspect: Evidence from Afrikaans. *Linguistics, 51*(5), 929–955.
- Bylund, E., & Jarvis, S. (2011). L2 effects on L1 event conceptualization. *Bilingualism: Language and Cognition, 14*, 47–59.
- Cadierno, T. (2004). Expressing motion events in a second language: A cognitive typological approach. In M. Achard, & S. Niemeier (Eds.), *Cognitive linguistics, second language acquisition and foreign language pedagogy* (pp. 13–49). Mouton de Gruyter.
- Cadierno, T., & Ruiz, L. (2006). Motion events in Spanish L2 acquisition. *Annual Review of Cognitive Linguistics, 4*, 183–216.
- Cardini, F. (2008). Manner of motion saliency: An inquiry into Italian. *Cognitive Linguistics, 19*(4), 533–570.
- Cardini, F. (2012). Grammatical constraints and verb-framed languages: The case of Italian. *Language and Cognition, 4*(3), 167–201.
- Carroll, M., & Lambert, M. (2006). Reorganizing principles of information structure in advanced L2s: A study of French and German learners of English. In H. Byrnes, H. Weger-Guntharp, & K. Sprang (Eds.), *Educating for Advanced Foreign Language Capacities* (pp. 54–73). Georgetown University Press.
- Carroll, M., Weimar, K., Flecken, M., Lambert, M., & von Stutterheim, C. (2012). Tracing trajectories: Motion event construal by advanced L2 French-English and L2 French-German speakers. *Language, Interaction and Acquisition, 3*(2), 202–230.
- Durst-Andersen, P., Smith, V., & Thomsen, O. N. (2013). Towards a cognitive-semiotic typology of motion verbs. In C. Paradis, J. Hudson, & U. Magnusson (Eds.), *The Construal of Spatial Meaning* (pp. 118–143). Oxford University Press.

- Ellis, N. C. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned-attention, and the limited L2 end-state. In P. Robinson, & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 372–405). Routledge.
- Feist, M. I. (2016). Minding your manners: linguistic relativity in motion. *Linguagem em (Dis)curso*, 16(3), 591–602.
- Flecken, M., von Stutterheim, C., & Carroll, M. (2014). Grammatical aspect influences motion event perception: Evidence from a cross-linguistic, non-verbal recognition task. *Language and Cognition*, 6, 45–78.
- Flecken, M., Carroll, M., Weimar, K., & von Stutterheim, C. (2015). Driving along the road or heading for the village? Conceptual differences underlying motion event encoding in French, German, and French–German L2 users. *The Modern Language Journal*, 99(S1), 100–122.
- Gerwien, J., & von Stutterheim, C. (2018). Event segmentation: Cross-linguistic differences in verbal and non-verbal tasks. *Cognition*, 180, 225–237.
- Gerwien, J., & Stutterheim, C. v. (2022). Describing motion events. In A. H. Jucker, & H. Hausendorf (Eds.), *Handbook of Pragmatics (HOPS) 14: The Pragmatics of Space* (pp. 152–180). De Gruyter Mouton.
- Hickmann, M., & Robert, S. (Eds.). (2006). *Space in languages: Linguistic systems and cognitive categories*. John Benjamins.
- Hickmann, M., Hendriks, H., & Champaud, C. (2008). Typological constraints on motion in French child language. In J. Guo, E. Lieven, S. Ervin-Tripp, N. Budwig, S. Özçalışkan, & K. Nakamura (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 209–224). Lawrence Erlbaum Associates.
- Jarvis, S. (2007). Theoretical and methodological issues in the investigation of conceptual transfer. *Vigo International Journal of Applied Linguistics*, 4, 43–71.
- Jarvis, S., & Pavlenko, A. (2008). *Cross-linguistic influence in language and cognition*. Routledge.
- Kartal, G., & Sarigul, E. (2017). Frequency effects in second language acquisition: An annotated survey. *Journal of Education and Training Studies*, 5(6), 1–8.
- Kopecka, A. (2009). L'expression du déplacement en français: L'interaction des facteurs sémantiques, aspectuels et pragmatiques dans la construction du sens spatial. *Langages*, 1(173), 54–75.
- Lambert, M., von Stutterheim, C., & Gerwien, J. (2022). Under the surface: a survey of principles of language use in advanced L2 speakers. *Language, Interaction and Acquisition*, 13(1), 1–28.
- Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity*. Cambridge University Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.
- Newtonson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28(1), 28–38.

- Özçalışkan, S., & Emerson, S. (2016). Learning to talk, think, and gesture about motion in language-specific ways: Insights from Turkish. In N. Ketrez, & B. Haznedar (Eds.), *Trends in language acquisition research: The acquisition of Turkish in Childhood* (pp. 177–191). John Benjamins.
- Papafragou, A., & Selimis, S. (2010). Event categorization and language: A cross-linguistic study of motion. *Language and Cognitive Processes*, 25, 224–260.
- Pavlenko, A. (2014). *The bilingual mind: And what it tells us about language and thought*. Cambridge University Press.
- Pavlenko, A., & Vaolynsky, M. (2015). Motion encoding in Russian and English: Moving beyond Talmy's typology. *Modern Language Journal*, 99 (special issue), 32–48.
- Pourcel, S. (2005). What makes path of motion salient? *Berkeley Linguistic Society (BLS)*, 30, 505–516.
- Pourcel, S., & Kopecka, A. (2005). Motion expression in French: Typological diversity. *Durham & Newcastle Working Papers in Linguistics*, 11, 139–153.
- Slobin, D. I. (1991). Learning to think for speaking: Native language, cognition and rhetorical style. *Pragmatics*, 1, 7–26.
- Slobin, D. I. (1996). From thought and language to thinking for speaking. In J. Gumperz, & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). University Press.
- Slobin, D. I. (2004). The many ways to search for a frog. Linguistic typology and the expression of motion events. In S. Strömquist, & L. Verhoeven (Eds.), *Relating events in narrative* (Vol. 2, pp. 219–257). Lawrence Erlbaum Associates.
- Slobin, D. I. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann, & S. Robert (Eds.), *Space in languages: Linguistic systems and cognitive categories* (pp. 59–81). John Benjamins.
- Slobin, D. I., Bowerman, M., Brown, P., Eisenbeiß, S., & Narasimhan, B. (2011). Putting things in places: Developmental consequences of linguistic typology. In J. Bohnemeyer, & E. Pederson (Eds.), *Event representation in language and cognition* (pp. 134–165). University Press.
- Stringer, D. (2005). Children's structuring of motion events: Syntactic universals and lexical variation. In Y. Ostu (Ed.), *Proceedings of the Sixth Tokyo Conference on Psycholinguistics* (pp. 319–343). Hitsuji Shobo.
- Talmy, L. (1975). Semantics and syntax of motion. In J. Kimball (Ed.), *Syntax and semantics 4*. Academic Press.
- Talmy, L. (2000). *Toward a cognitive semantics*. Cambridge MIT Press.
- Treffers-Daller, J., & Tidball, F. (2015). Can L2 learners learn new ways to conceptualize events? Evidence from motion event construal among English-speaking learners of French. In P. Guijarro-Fuentes, K. Schmitz, & N. Müller (Eds.), *The acquisition of French in multi-lingual contexts* (pp. 145–184). Multilingual Matters.

- von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtova, B. (2012). How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics*, 50(4), 833–867.
- von Stutterheim, C., Bouhaous, A., & Carroll, M. (2017). From time to space: The impact of aspectual categories on the construal of motion events: The case of Tunisian Arabic and Modern Standard Arabic. *Linguistics*, 55(1), 207–249.
- von Stutterheim, C., & Carroll, M. (2006). The impact of grammatical temporal categories on ultimate attainment in L2 learning. In H. Byrnes, H. Weger-Guntharp, & K. Sprang (Eds.), *Educating for Advanced Foreign Language Capacities* (pp. 40–53). Georgetown University Press.
- von Stutterheim, C., Carroll, M., & Klein, W. (2009). New perspectives in analysing aspectual distinctions across languages. In W. Klein, & P. Li (Eds.), *The expression of time* (pp. 195–216). Mouton de Gruyter.
- von Stutterheim, C., Flecken, M., & Carroll, M. (Eds.). (2013). *Principles of information organization in language use: On the L2 acquisition of complex conceptual structures* (IRAL Special Issue). De Gruyter.
- von Stutterheim, C., Gerwien, J., Bouhaous, A., Carroll, M., & Lambert, M. (2020). What makes up a reportable event in a language? Motion events as an important test domain in linguistic typology, *Linguistics*, 58(6), 1659–1700.
- von Stutterheim, C., Lambert, M., & Gerwien, J. (2021). Limitations on the role of frequency in L2 acquisition. *Language and Cognition*, 13(2), 291–321.
- von Stutterheim, C., & Nüse, R. (2003). Processes of conceptualisation in language production. *Linguistics*, 41(5), 851–881.
- Wulf, S., & Ellis, N. C. (2018). Usage-based approaches to second language acquisition. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual Cognition and Language: The state of the science across its subfields* (pp. 37–56). John Benjamins.

---

Hana Klages

# „Es ist *der* Ritter, obwohl du ‚*sie*‘ gesagt hast“: Auflösung anaphorischer Personalpronomina im Erwerb des Deutschen als Erst- und Zweitsprache

---

**Zusammenfassung** In diesem Artikel werden Ergebnisse aus einer Entscheidungsaufgabe präsentiert, mit der das Verstehen anaphorischer Personalpronomina durch 6- und 7-jährige Kinder mit Deutsch als Erstsprache (DaE) und 5-, 6-, und 7-jährige Kinder mit Deutsch als Zweitsprache (DaZ) untersucht wurde. Die Ergebnisse zeigen, dass sich DaE- und DaZ-Kinder während der Pronomenauflösung unterscheiden. Während Kinder mit DaE spätestens ab dem 6. Lebensjahr die beiden untersuchten Cues *Genus-NGP* und *Cue Konzept* während der Pronomenauflösung zuverlässig verarbeiten und koordinieren, ist diese Fähigkeit bei einigen Kindern mit DaZ noch nicht zuverlässig erworben. Des Weiteren legen die Befunde nahe, dass die Gruppe der DaZ-Kinder sich weitaus heterogener verhält als die DaE-Gruppe. So fanden sich bereits in der Gruppe der 5-jährigen DaZ-Kinder Individuen, die während der Pronomenauflösung die Cues *Genus NGP* und *Konzept* verarbeiten und koordinieren. Gleichzeitig waren in der Gruppe der 7-Jährigen immer noch Kinder vertreten, die diese beiden Cues noch nicht berücksichtigen und stattdessen auf andere, vermutlich frühere Auflösungsstrategien zurückgreifen. Hierzu gehören Auflösungsstrategien, bei denen die Position bzw. syntaktische Rolle des Antezedenten eine Rolle spielen. Für die DaZ-Sprachdidaktik bedeuten unsere Ergebnisse, dass die Förderung der Fähigkeiten zur Auflösung anaphorischer Personalpronomina ein unabdingbarer Bestandteil sprachdidaktischer Interventionen sein sollte.

**Schlüsselwörter** Deutsch als Zweitsprache; Anaphernaauflösung; Spracherwerb; Sprachrezeption; Pronomen

## 1 Einleitung

Mit dem im Titel enthaltenen Satz „Es ist der Ritter, obwohl du *sie* gesagt hast.“ hat mir einmal ein 6-jähriges Kind aus der Nachbarschaft geantwortet, nachdem ich ihm folgenden Dreizeiler vorgelesen hatte:

Die Freunde spielen Verstecken.  
Die Hexe sucht den Ritter.  
Sie versteckt sich hinter einem Vorhang.

Die Antwort zeigt, welche komplexen Gedanken sich der kleine Junge gemacht hat, als er diesen einfachen Text verarbeitet hat. Er wägte ab, ob der Referent des Subjektpronomens „*sie*“ im dritten Satz auf der Grundlage der *Genuskongruenz* zwischen Pronomen und einem der vorausgegangenen Ausdrücke identifiziert werden kann, oder ob der *Kontext*, in dem das Pronomen eingebettet war, ausschlaggebend für die Personenidentifizierung ist. Die Antwort zeigt nun, dass der Junge sich schließlich für den *Kontext* entschieden hat, ihm jedoch wichtig war, mir mitzuteilen, dass er sehr wohl auch um die Rolle der Genuskongruenz bei der Auflösung anaphorischer Personalpronomina weiß.

Mein Interesse wurde geweckt und ich fragte mich, wie wohl erwachsene SprecherInnen des Deutschen oder ein Kind, das Deutsch als Zweitsprache erwirbt, geantwortet hätten. Dieser Gedanke ließ mich nicht los und ich beschloss, dieser Frage nachzugehen. Ich fing an, nach Studien zur Auflösung anaphorischer Personalpronomina durch verschiedene SprecherInnen zu suchen, und tauchte in ein Meer an Publikationen ein, die sich mit dieser Thematik befassen. Bald standen die Fragestellung und das Untersuchungsdesign für eine eigene Eye-Tracking-Studie zur Rolle verschiedener Auflösungs hinweise während der Textrezeption durch 5- bis 9-jährige Kinder mit Deutsch als Erstsprache (DaE) und Deutsch als Zweitsprache (DaZ) fest. Bevor ich mit der Datenerhebung begonnen habe, wollte ich erfahren, ob die angedachten Texte und Bilder von den Kindern erkannt und verstanden werden. Herausfinden wollte ich es im Rahmen der vorliegenden Studie.

Nachdem ich das vorgesehene Untersuchungsmaterial bei Kindern mit DaE und DaZ aus dem Freundes- und Bekanntenkreis eingesetzt hatte, wurde schnell klar, dass zwar einige kleine Veränderungen vorgenommen werden müssen, das Testmaterial aber in Gänze geeignet ist.

Die kindlichen Antworten, die ich im Rahmen dieser Studie notiert habe, habe ich zunächst auf die Seite gelegt. Erst einige Zeit später nahm ich meine Notizen wieder in die Hand und schaute genauer hin. Ich entdeckte in den kindlichen Antworten Regularitäten, die mit dem Alter (5-, 6-, und 7-Jährige), dem Erwerbskontext der Kinder (DaE und DaZ) und dem Texttyp (zu den verwendeten Texttypen s. Kap. 3) korrelierten.

In diesem Beitrag sollen die Studie und ihre Ergebnisse präsentiert werden. In Kapitel 2 werden zunächst die verschiedenen sprachlichen Auflösungsclues vorgestellt, also die Informationen, die beim Verstehen anaphorischer Personalpronomina eine Rolle spielen. Danach werden einige Erkenntnisse zur Verarbeitung dieser Clues in der Erstsprache (L1) und der Zweitsprache (L2) insbesondere bei Kindern referiert und anschließend eine Forschungsfrage formuliert, die im Rahmen dieser Studie beantwortet wurde. Das Untersuchungsdesign sowie die Ergebnisse der Studie werden in den Kapiteln 3 und 4 präsentiert. Der Beitrag schließt mit Kapitel 5 ab, in dem die Ergebnisse zusammengefasst werden.<sup>1</sup>

## 2 Verarbeitung anaphorischer Personalpronomina

Während der Auflösung anaphorischer Personalpronomina in Texten stehen die RezipientInnen vor der Aufgabe, den passenden Diskursreferenten zu finden, also diejenige Entität aus der bereits erzeugten mentalen Diskursrepräsentation zu identifizieren, auf die das anaphorische Pronomen referiert. Dieser Prozess der Referenzzuordnung zu einem anaphorischen Pronomen wird *Pronomenauflösung* genannt. Für die Pronomenauflösung stehen den RezipientInnen eine Reihe verschiedener sprachlicher Informationen zur Verfügung. Diese Informationen werden Auflösungshinweise bzw. Auflösungsclues genannt (Klages & Gerwien, 2015; Klages & Gerwien, 2020). Hierzu gehören zum Beispiel Genus und Numerus des Personalpronomens und der vorausgegangenen Ausdrücke (vgl. Crawley et al., 1990; Ehrlich, 1980; Shillcock, 1982; Arnold et al., 2000) oder der Kontext, in dem das Pronomen eingebettet ist (vgl. McKoon & Ratcliff, 1992). Auch der Topikstatus der vorausgegangenen Ausdrücke (Topik vs. Non-Topik; vgl. Bosch & Umbach, 2007), die *syntaktische Rolle* der vorausgegangenen Ausdrücke (Subjekt vs. nicht Subjekt; vgl. Järvikivi et al., 2005; Frederiksen, 1981) oder ihre *Nennungsposition* im Satz (Erstnennung vs. Nicht-Erstnennung; vgl. Gernsbacher & Hargreaves, 1988; Arnold et al., 2007) spielen bei der Pronomenauflösung eine Rolle. Für das Deutsche gilt, dass anaphorische Personalpronomina bevorzugt auf den Referenten im Topik Bezug nehmen, auf den im vorausgegangenen Satz meist mit dem Ausdruck in Subjekt bzw. in Erstnennung referiert wurde (vgl. Ellert, 2011; Bryant & Noschka, 2015).

Beispiele 1 bis 4 illustrieren Texte mit verschiedenen Clues: Beispiel 1 enthält den Cue *Genus*, Beispiel 2 den Cue *Numerus*, Beispiel 3 den Cue

---

1 Ich danke Ines Marberg, Christoph Jörger und Ingeborg Gutfleisch für wertvolle Kommentare, Hinweise und Anregungen.

*Einbettungskontext* (hier im Folgenden *Konzept* genannt) und Beispiel 4 den Cue *syntaktische Rolle* bzw. *Nennungsposition*:

- (1) **Ingo** hat Henriette besucht. Er hat einen Blumenstrauß gekauft.
- (2) **Margarete** hat ihre Freunde besucht. Sie hat einen Blumenstrauß gekauft.
- (3) Henriette hat **ihre Mutter** vom Bahnhof mit dem Auto abgeholt. Sie hatte nämlich viel Gepäck und eine Busfahrt wäre nicht zumutbar gewesen.
- (4) **Henriette** fährt mit Margarete in den Urlaub. Sie freut sich schon auf die gemeinsame Zeit.

In einem Diskurs stehen den RezipientInnen in der Regel mehrere Auflöscues zur Verfügung. Die Aufgabe besteht darin, die verschiedenen Cues zu erkennen und zu koordinieren. Im Rahmen der Constraint-based-Ansätze (z. B. Tannenhaus & Trueswell, 1995) wird angenommen, dass für die Cue-Koordination ein Sprachverarbeitungsmechanismus zuständig ist, der inkrementell die eingehenden multiplen (syntaktischen, grammatischen, konzeptionellen etc.) Informationen aus dem Input und aus dem Weltwissen in parallel ablaufenden Prozessen verarbeitet. Die Ergebnisse der Verarbeitung, das heißt die syntaktischen, lexikalischen und diskursiven mentalen Repräsentationen, können dabei zu bestimmten Zeitpunkten mit der inkrementellen Verarbeitung inkompatibel sein (s. Beispiel 5).

- (5) Die Hexe sucht den Ritter. Sie versteckt sich hinter einem Vorhang.

Im Beispiel 5 signalisiert der Cue Genus, dass das Pronomen auf den Referenten, auf den im vorausgegangenen Satz mit dem Ausdruck „Hexe“ Bezug genommen wird, rekurriert. Dieser Ausdruck wird als Antezedent des anaphorischen Personalpronomens bezeichnet (Schwarz-Friesel, 2000). Dagegen suggeriert der Kontext, in dem das Pronomen eingebettet ist, gepaart mit dem Welt- und Konzeptwissen der RezipientInnen, dass hinter dem Vorhang „der Ritter“ steht, er also als der Referent des Personalpronomens „sie“ gilt. Welcher Cue im Prozess der Rezeption höher gewichtet wird, hängt nach Annahmen der Cue-Reliability-Ansätze von dem Grad ihrer Reliabilität (Arnold et al., 2005; Arnold et al., 2007) ab. Unter Reliabilität wird die Verlässlichkeit verstanden, mit der ein Cue zur Identifikation des passenden Referenten beiträgt (Klages & Gerwien, 2015). Zu den Cues mit einer hohen Reliabilität oder anders ausgedrückt, zu den reliablen Cues, zählt in Sprachen, die über die Kategorie Genus verfügen, die Genuskongruenz zwischen Pronomen und dem Antezedenten. Da Verstöße gegen die Genuskongruenz in Genus-Sprachen so gut wie gar nicht vorkommen, so Klages und Gerwien (2015,



S. 76), „gilt das Genus in der internen Grammatik kompetenter Sprecher als hochgradig reliabler und für die Referentenbestimmung maßgeblicher Hinweis.“ Zu den weniger reliablen, das heißt, zu den probabilistischen Cues, zählt nach Annahme der Cue-Reliability-Ansätze zum Beispiel die *syntaktische Rolle* oder *Nennungsposition* der vorausgegangenen Nominalphrasen. Die probabilistischen Cues sind den reliablen Hinweisen während der Verarbeitung untergeordnet.

Kompetenten SprecherInnen gelingt die Koordination verschiedener Cues schnell und mit einem geringen kognitiven Aufwand (Arnold et al., 2007; Arnold et al., 2000). Lernende einer Sprache als L1 (also Kinder) oder als L2 (also Kinder oder Erwachsene) müssen das Wissen über die spezifischen Cues in der Zielsprache erst erwerben. Kinder stehen außerdem noch vor der Aufgabe, die generelle Fähigkeit zur Cue-Koordination und Cue-Gewichtung zu erwerben.

Erkenntnisse zum Erwerb der Fähigkeit zur Auflösung anaphorischer Personalpronomina liegen für verschiedene Sprachen und Erwerbstypen vor. Für den Erwerb der Erstsprache geht man etwa im Rahmen der Cue-Reliability-Ansätze davon aus, dass der Erwerb der Prozessierung von reliablen Cues (z. B. Genus) früher erfolgt als der Erwerb der probabilistischen Cues (Arnold et al., 2005; Arnold et al., 2007). Empirische Evidenz für diese Annahme wurde im Rahmen von Online- als auch Offline-Aufgaben beispielsweise für den Erwerb des Englischen als L1 erbracht Arnold et al., 2007. Im Rahmen der First-Mention-Prominence-Ansätze (Song & Fisher, 2005, 2007) wird dagegen von dem Primat der *Nennungsposition* der potenziellen Antezedenten während des Erwerbs der L1 ausgegangen. Diesem Ansatz zufolge berücksichtigen Kinder bereits im Alter von drei bis vier Jahren die Nennungsabfolge im Diskurs und nutzen sie, um anaphorische Pronomina aufzulösen. Empirische Evidenz für die Berücksichtigung der syntaktischen Rolle der Referenzausdrücke während der Auflösung anaphorischer Personalpronomina durch Kinder wurde von Song & Fisher (2005) für den Erwerb des Englischen als L1, für den Erwerb des Russischen als L1 von Gagarina (2007) und für den Erwerb des Deutschen als L1 beispielsweise von Järvikivi et al. (2014), von Bittner (2007) und von Klages & Gerwien (2015) erbracht.

Für den Erwerb der Auflösung anaphorischer Personalpronomina in der L2 ist die Forschungslandschaft durch zwei Ansätze geprägt: die Shallow-Structure-Hypothese (z. B. Clahsen & Felser, 2006, 2018) und die Interface-Hypothese (z. B. Sorace & Filiaci, 2006). Der Shallow-Structure-Hypothese zufolge spielen während der Sprachrezeption einer L2 primär semantische (z. B. Synonymie), konzeptuelle (z. B. Sinnhaftigkeit des Textes oder Weltwissen) und pragmatische Cues (z. B. Topikalität) eine Rolle, die grammatischen Cues wie auch Genus werden eher weniger stark berücksichtigt. Die Interface-Hypothese geht davon aus, dass die Koordination und Integration

multipler Cues wie der pragmatischen, semantischen und grammatischen Cues in der L2 den RezipientInnen Schwierigkeiten bereitet, was dazu führt, dass die Sprachverarbeitung primär unter Berücksichtigung der grammatischen Informationen erfolgt. Empirische Evidenz liegt für beide Hypothesen vor (Pan et al., 2015; Sorace & Filiaci, 2006).

Für den Erwerb des Deutschen als Zweitsprache durch Kinder liegen vergleichsweise wenige Studien vor. Der Erwerb probabilistischer Hinweise wie *Nennungsposition* und Topikalität wurde von Bryant & Noschka (2015) untersucht. Die AutorInnen fanden heraus, dass der Erwerb dieser Cues durch Kinder mit DaZ später abgeschlossen ist als bei Kindern mit DaE und mindestens bis zum 10. Lebensjahr andauert. Klages & Gerwien (2015) und Klages & Gerwien (2020) untersuchten den Erwerb und die Koordination von reliablen Cues (*Genus NGP*<sup>2</sup>) und probabilistischen Cues (hier: *syntaktische Rolle / Nennungsposition*). Die AutorInnen fanden heraus, dass bereits 5-jährige Kinder beide Cues systematisch berücksichtigen und koordinieren, das Ausmaß der Berücksichtigung des reliablen Cues *Genus NGP* bei DaZ-Kindern in allen Altersgruppen jedoch unter dem Ausmaß der Berücksichtigung bei Kindern mit DaE liegt. Mit Blick auf den Erwerb des Cues *syntaktische Rolle / Nennungsposition* stellen die AutorInnen im untersuchten Altersspektrum von fünf bis neun Jahren eine Verschiebung der Verarbeitungspräferenzen fest: 5-Jährige lösen anaphorische Personalpronomina in Richtung des vorausgegangenen Objekts / Zweitnennung auf, während die 9-Jährigen das Subjekt / die Erstnennung des vorausgegangenen Satzes als den präferierten Antezedenten wählen. Außerdem fanden die AutorInnen heraus, dass die Prozessierung und Koordination beider Cues bis zum 9. Lebensjahr zunehmend automatisiert wird.

Zur Rolle des Kontextes während der Anaphernauflösung im L1- und L2-Erwerb des Deutschen oder anderer Sprachen durch Kinder liegen meines Wissens keine neueren empirischen Studien vor. Die vorhandene Forschungslücke soll mit Erkenntnissen aus der vorliegenden Studie, in der diese beiden Cue-Typen im Fokus stehen, gefüllt werden. Folgende Forschungsfrage soll in diesem Rahmen beantwortet werden: Welche Präferenzen lassen sich mit Blick auf die Verarbeitung des Cues *Genus NGP* und des Cues *Konzept* bei Kindern im Alter zwischen fünf und sieben Jahren im Erwerb von DaE und DaZ feststellen?

---

2 NGP steht für natürliches Geschlechtsprinzip. Dem NGP zufolge erhalten Nomina ihr grammatisches Genus entsprechend dem Sexus des Referenten, auf den sie sich beziehen. So ist z. B. das Substantiv *Frau* feminin, da es auf eine weibliche Person referiert und das Substantiv *Onkel* maskulin, da es auf eine männliche Person referiert.

### 3 Untersuchungsdesign

Zur Beantwortung der obigen Forschungsfrage wurden Daten aus einer Entscheidungsaufgabe ausgewertet. Bei dieser Aufgabe entschieden die Untersuchungsteilnehmenden, welches von zwei Bildern zu einer auditiv präsentierten Kurzgeschichte passt. Wie bereits erwähnt, wurde die Studie primär als eine Pilotstudie zu einer umfassenderen Untersuchung zur Online-Auflösung anaphorischer Personalpronomina durch Kinder mit Deutsch als Erst- und Zweitsprache konzipiert (Klages & Gerwien, 2015; Klages & Gerwien, 2020) und diente primär der Überprüfung der Eignung der verwendeten Stimuli (Bilder und Texte).

#### Teilnehmende

Ausgewertet wurden Daten von 24 monolingual deutschsprachigen ( $n=10$ ) und zweisprachigen ( $n=14$ ) Kindern, die neben dem Deutschen als Zweitsprache bereits eine andere Sprache als Erstsprache erworben hatten. Die Größe und Zusammensetzung der Stichprobe ist dem primär explorativen Charakter der Studie und den verfolgten Zielen geschuldet. Alle Kinder besuchten zum Zeitpunkt der Datenerhebung eine Bildungseinrichtung mit Deutsch als Kommunikationssprache (KiTa oder Grundschule). Die mehrsprachigen Kinder hatten außerdem einen regelmäßigen Kontakt zu der Erstsprache. Der Kontakt zum Deutschen setzte bei den Kindern mit DaZ mit dem Eintritt in den Kindergarten ein, also mit ca. 3 Jahren. Aufgeteilt nach Alter (5-Jährige, 6-Jährige und 7-Jährige) und Erwerbstyp (DaE und DaZ) ergaben sich fünf Untersuchungsgruppen (s. Tabelle 1).

Tabelle 1. Überblick über die Untersuchungsteilnehmenden

	DaE	DaZ
5-Jährige	–	$n=5$
6-Jährige	$n=5$	$n=5$
7-Jährige	$n=5$	$n=4$

#### Material

Für die Studie wurden 18 Texte konstruiert, davon 16 Testitems und 2 Kontrollitems. Jeder Text (sowohl Test- als auch Kontrollitems) bestand aus drei miteinander unverbundenen Hauptsätzen: einem Einführungssatz, einem

Antezedentsatz und einem Testsatz. Die Texte wurden während des Experiments von der Leiterin vorgetragen.

Um die Verarbeitung der Auflösungsclues *Genus NGP* und *Konzept* untersuchen zu können, wurden 4 Typen von Texten kreiert. In den Typen 1 und 2 ist jeweils einer der beiden Auflösungsclues enthalten: im Typ 1 *Genus NGP* und im Typ 2 *Konzept*. Das Pronomen im Testsatz bezieht sich zu 50 % auf das Subjekt und zu 50 % auf das Objekt im Antezedentsatz, feminine und maskuline Pronomina sind jeweils zu 50 % vertreten.

Typ 1: ein Cue *Genus NGP* (G)

Einführungssatz: Die Leute spielen Verstecken.  
 Antezedentsatz: Die Hexe sucht den Koch.  
 Testsatz: Sie/Er hat neue Schuhe an.

Typ 2: ein Cue *Konzept*<sup>3</sup> (K)

Einführungssatz: Die Leute spielen Verstecken.  
 Antezedentsatz: Die Krankenschwester sucht die Hexe.  
 Testsatz: Sie zählt zunächst bis 10. / Sie versteckt sich hinter einem Vorhang.

Texte vom Typ 3 und 4 enthalten zwei Auflösungshinweise, sowohl *Genus NGP* als auch den *konzeptuellen* Hinweis. Im Typ 3 zeigen beide Cues auf ein und denselben Antezedenten (kongruente Bedingung), im Typ 4 weist jeder Cue auf einen anderen potenziellen Antezedenten hin (inkongruente Bedingung).

Typ 3: zwei Cues – kongruent (GK)

Einführungssatz: Die Leute spielen Verstecken.  
 Antezedentsatz: Die Hexe sucht den Ritter.  
 Testsatz: Er versteckt sich hinter einem Vorhang. /  
 Sie zählt zuerst bis zehn.

---

3 In den Testitems von Typ 1 und Typ 2 wurden in den Testsätzen die Pronomen *er* und *sie* jeweils zwei Mal verwendet. Sie beziehen sich jeweils einmal auf das Subjekt und einmal auf das Objekt des vorausgegangenen Satzes. Die im Antezedentsatz verwendeten Substantive gehören im Typ 1 zwei unterschiedlichen Genusklassen an (feminin und maskulin), im Typ 2 dagegen ein und derselben Genusklasse (entweder feminin oder maskulin).

#### Typ 4: zwei Cues – inkongruent<sup>4</sup> (GxK)

Einführungssatz: Die Leute spielen Verstecken.  
Antezedentsatz: Die Hexe sucht den Ritter.  
Testsatz: Sie versteckt sich hinter einem Vorhang. /  
Er zählt zuerst bis zehn.

Die visuellen Stimuli bestanden aus farbigen Abbildungen von den in den Texten genannten Figuren (s. Abbildung 1).



Abbildung 1. Beispiele visueller Stimuli

Die Kontrollitems haben die gleiche Struktur wie die Testitems, enthalten aber weder den Cue *Genus NGP* noch den Cue *Konzept*.

Kontrollitems:

Einführungssatz: Die Leute spielen Verstecken.  
Antezedentsatz: Die Königin sucht die Hexe.  
Testsatz: Sie hat braune Haare.

---

4 Die Häufigkeit der Verwendung der Pronomen im Testsatz in den Typen 3 und 4 ist identisch mit der Häufigkeit in den Typen 1 und 2: Jedes Pronomen (*er* und *sie*) kommt zwei Mal vor. Im Typ 3 (kongruente Bedingung) bezieht sich das Pronomen so wie bei den Typen 1 und 2 einmal auf das Subjekt und einmal auf das Objekt des vorausgegangenen Satzes. Im Typ 4 (inkongruente Bedingung), in dem jeder der beiden Cues auf einen anderen Referenz Ausdruck des vorausgegangenen Satzes Bezug nimmt, zeigt sowohl der Cue *Genus NGP* als auch der Cue *Konzept* einmal auf das Subjekt und einmal auf das Objekt.

## Untersuchungsverlauf

Die Datenerhebung erfolgte in einer Sitzung von maximal 15 Minuten. Während der Sitzung saßen die Kinder und die Experimentleiterin an einem Tisch. Die Kinder wurden aufgefordert, sich kurze Geschichten anzuhören und von zwei Bildern dasjenige Bild auszuwählen (per Zeigegeste), das zum letzten Satz der Geschichte passt. Die zwei Bilder wurden vor Beginn jeder Geschichte vor die Kinder auf den Tisch gelegt. Um sicherzugehen, dass die Kinder alle verwendeten Figurenbezeichnungen kennen, sollten sie vor Beginn des Experiments alle abgebildeten Figuren benennen. Die Entscheidungen der Kinder wurden während des Experiments schriftlich notiert. Gleichzeitig wurde das gesamte Experiment aufgenommen, um evtl. Rückfragen oder Kommentare der Kinder festzuhalten.

## Analyse

Die gewonnenen Daten wurden mit zwei Werten kodiert: mit den Werten „korrekt“ und „inkorrekt“. Bei den Items vom Typ 1 (G) und 2 (K) wurde eine Entscheidung dann mit „korrekt“ kodiert, wenn die Pronomenauflösung auf der Grundlage des *Genus-Cues* (Typ 1) oder des *Cues Konzept* (Typ 2) erfolgte. Bei den Items vom Typ 3 (GK), in dem die beiden Cues auf ein und denselben Antezedenten hindeuten, wurde eine Entscheidung mit „korrekt“ versehen, wenn ebendieser Antezedent ausgewählt wurde. Im Typ 4 (GxK) wurde eine Entscheidung mit „korrekt“ versehen, wenn die Pronomen auf der Grundlage der Genuskongruenz zwischen Pronomen und dem vorausgegangenen Substantiv aufgelöst wurden. Entscheidungen bei Typ (GxK), die auf der Grundlage des *Cues Konzept* getroffen wurden, wurden also mit „inkorrekt“ kodiert.

Anschließend wurde eine deskriptive statistische Analyse mit den unabhängigen Variablen Gruppe (DaZ vs. DaE), Alter (5- vs. 6- vs. 7-Jährige), und Bedingung (G, K, GK, GxK) durchgeführt. Auf eine inferenzstatistische Auswertung wird hier aufgrund der geringen Stichprobengröße verzichtet. Die im Kap. 4 formulierten Aussagen sind also primär als erste, nicht generalisierbare Erkenntnisse zu verstehen. Sie stellen erste Anhaltspunkte zur Hypothesenformulierung dar, die im Rahmen einer umfassenderen und gezielten Untersuchung zu überprüfen wären.

## 4 Ergebnisse und Diskussion

Im Rahmen der Studie wurden insgesamt 384 Datenpunkte erhoben. 41,67% davon (n=160) wurden in der DaE-Gruppe elizitiert, 58,33% (n=224) in der DaZ-Gruppe. Die etwas ungleich verteilten Anteile der erhobenen

Datenpunkte zwischen den beiden Erwerbsgruppen sind insbesondere auf das Fehlen von 5-jährigen DaE-Kindern zurückzuführen.

Die gewonnenen Ergebnisse werden getrennt pro Gruppe vorgestellt und anschließend zusammengefasst.

### Pronomenauflösung DaE

Die Anteile der korrekten Antworten (in Prozent) für die Gruppe der DaE-Kinder aufgeteilt nach Bedingung und Altersgruppe sind graphisch in der Tabelle 2 dargestellt.

Tabelle 2. Anteile korrekter Antworten in Abhängigkeit von Alter und Bedingung bei Kindern mit DaE (in %)

	6-Jährige	7-Jährige	alle
G	85	95	90
K	85	100	92,5
GK	80	100	90
G x K	55	50	52,5
G, K, GK	83,3	98,3	90,8
alle	76,3	86	81,3

Zu den Hauptbefunden zählt, dass Kinder mit DaE anaphorische Personalpronomina in den untersuchten Kontexten erfolgreich auflösen: Sie identifizieren bei 81,3% der Stimuli den korrekten Antezedenten. Der Anteil der korrekten Antworten fällt noch höher aus, wenn man lediglich die eindeutigen Items berücksichtigt (Typ G, K, und GK). Hier beträgt der Anteil der korrekten Antworten 90,8%.

Was die Effekte der Bedingung anbelangt, so konnten zwischen den zwei Ein-Cue-Bedingungen (G und K) keine Unterschiede festgestellt werden, weder bei der Betrachtung aller Kinder noch bei der Betrachtung einzelner Altersgruppen (s. Tabelle 2). Für die zwei unterschiedlichen Zwei-Cue-Bedingungen (GK und GxK) sind die Unterschiede dagegen deutlich, sowohl für alle Kinder als auch für die einzelnen Altersgruppen. Die Leistungen in der inkongruenten Zwei-Cue-Bedingung sind dabei wesentlich niedriger als in der kongruenten Zwei-Cue-Bedingung (s. Tabelle 2). Die Werte aus der kongruenten Zwei-Cue-Bedingung liegen in etwa auf dem Niveau der Leistungen aus den Ein-Cue-Bedingungen. Für die inkongruente Zwei-Cue-Bedingung bedeutet das jedoch nicht, dass die Kinder die anaphorischen Pronomina „inkorrekt“ auflösen.

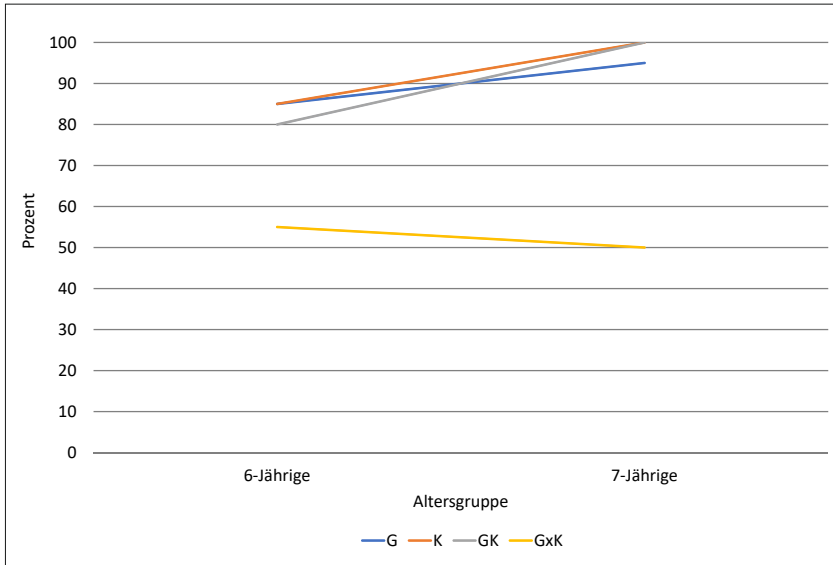


Abbildung 2. Anteile korrekter Antworten in allen Bedingungen (G), (K), (GK) und (GxK) in Abhängigkeit vom Alter bei Kindern mit DaE (in %)

Es bedeutet lediglich, dass einige der Antworten aus der Befolgung des Cues *Konzept* und nicht etwa des Cues *Genus NGP* resultieren. Zur dezidierten Diskussion der Leistungen in dieser Bedingung komme ich später. Zusammenfassend lassen sich die Befunde aus den kongruenten Bedingungen (G, K, GK) als Evidenz dafür deuten, dass Kinder mit DaE spätestens ab dem 6. Lebensjahr sowohl das *semantische Genus* (NGP) als auch den *Cue Konzept* systematisch für die Auflösung anaphorischer Personalpronomina nutzen.

Was die Effekte des Alters anbelangt, so konnte für die Ein-Cue-Bedingungen (G und K) eine positive Korrelation der Leistungen mit dem Alter beobachtet werden (s. Abbildung 2). Betrachtet man die Leistungen als Hinweis auf den Grad der Automatisierung der Verarbeitung und/oder als Hinweis auf den Grad der internen Validität dieser Cues, so lassen unsere Befunde auf einen Anstieg dieser Größen zwischen dem 6. und 7. Lebensjahr schließen. Vergleichbares wurde bei der Analyse der kongruenten Zwei-Cue-Bedingung gefunden. Auch hier lösen bereits die 6-jährigen Kinder das Pronomen überzufällig korrekt auf (80 %), die 7-jährigen sogar gänzlich zielsprachlich (s. Tabelle 2 und Abbildung 2). Für die inkongruente Zwei-Cue-Bedingung ist keine positive Korrelation zwischen Leistungen und Alter festzustellen, die Werte sind in den beiden Altersgruppen vergleichbar hoch und bewegen sich um das Zufallsniveau (s. Tabelle 2 und Abbildung 2).

Ob die Kinder in kongruenten Texten mit zwei Cues nur einen der beiden Hinweise berücksichtigen, also heuristisch vorgehen, oder beide Cues koordinieren und feststellen, dass sie dasselbe Ergebnis anzeigen, kann im Rahmen



der vorliegenden Studie nicht eindeutig beantwortet werden – mit beiden Strategien erzielen die Kinder das gleiche Ergebnis. Erst weitere Studien, die zum Beispiel ausschließlich und gezielt den Einfluss des Typs und der Anzahl der Auflösungs cues untersuchen, würden zur Klärung der Frage beitragen. Der leichte Abfall der Leistungen in der kongruenten Zwei-Cue-Bedingung gegenüber den Ein-Cue-Bedingungen bei den 6-Jährigen spricht aus meiner Sicht jedoch für die Annahme der Koordination der beiden Cues: Auflösungsprozesse, bei denen mehrere Cues wahrgenommen, ausgewertet und abgeglichen werden müssen, sind insbesondere in Stadien, in denen eine derartige Verarbeitung noch nicht ausreichend automatisiert ist, nicht nur ressourcenaufwendiger, sondern auch fehleranfälliger als eine Verarbeitung, bei der nur ein Cue wahrgenommen und ausgewertet wird (Klages & Gerwien, 2015). Der leichte Abfall der Leistungen in der kongruenten Zwei-Cue-Bedingung (80 % korrekte Antworten) bei den 6-Jährigen gegenüber den Ein-Cue-Bedingungen (85 % korrekte Antworten; s. Tabelle 2) könnte genau das widerspiegeln.

Weitere Argumente für diese Annahme liefern auch Befunde aus bisheriger Forschung. So gehen etwa Klages & Gerwien (2015, S. 85) auf der Grundlage der Ergebnisse aus einer Visual-World-Studie mit 5-, 7- und 9-jährigen Kindern mit DaE und DaZ davon aus, dass Kinder mit DaE im Alter von 7 Jahren verschiedene Auflösungs cues (*Genus NGP* und *syntaktische Rolle/Nennungsposition* des Antezedenten im vorausgegangenen Satz) koordinieren. Diese Fähigkeit wird der Studie zufolge im Erwerb von DaE bereits ab dem 5. Lebensjahr beobachtet und zwischen dem 5. und dem 7. Lebensjahr weiter ausgebaut.

Betrachten wir nun die Ergebnisse aus der inkongruenten Zwei-Cue-Bedingung (GxK). Einer rein quantitativen Analyse zufolge liegen hier die Antworten sowohl für die DaE-Gesamtgruppe als auch für die beiden Altersgruppen auf dem Zufallsniveau (55 % für die 6-Jährigen, 50 % für die 7-Jährigen, 52,5 % für alle Kinder). Das kann zum einen bedeuten, dass jedes Kind zufällig vorgeht und die Pronomina mal unter Berücksichtigung des Cues *Genus NGP* und mal unter Beachtung des Cues *Konzept* auflöst. Eine andere mögliche Erklärung für diese Befunde könnte in der individuellen Präferenz der einzelnen Kinder für einen der beiden Auflösungs cues liegen: Es ist denkbar, dass etwa die Hälfte den Kindern systematisch den einen und die andere Hälfte der Kinder systematisch den anderen Cue während der Verarbeitung präferiert.

Bei der Betrachtung der Antworten von einzelnen Kindern (Abbildung 3) wurde deutlich, dass die Kinder individuell tatsächlich einen der beiden Cues präferieren: Sowohl ein Teil der 6-Jährigen als auch ein Teil der 7-Jährigen folgen in der inkongruenten Bedingung überzufällig häufig<sup>5</sup> dem Cue *Genus NGP*. Die anderen Kinder in jeder Altersgruppe lösen die Pronomina dagegen

---

5 Unter ‚überzufällig häufig‘ verstehe ich hier, dass in mindestens 3 von 4 Antworten (75 %) ein und derselbe Cue verfolgt wird.

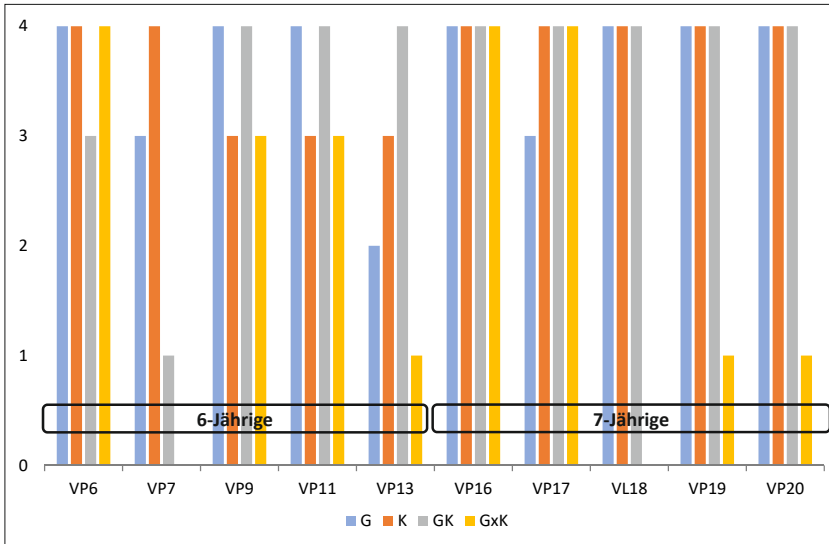


Abbildung 3. Korrekte Antworten für jedes Kind mit DaE (absolute Zahlen)

überzufällig häufig auf der Grundlage des Cues *Konzept* auf. Die individuelle Cue-Präferenz in der inkongruenten Bedingung korreliert dabei interessanterweise bei den 6-Jährigen mit der höheren Anzahl der korrekten Antworten in der entsprechenden Ein-Cue-Bedingung: Folgt z. B. ein Kind in der inkongruenten Bedingung dem Cue *Genus NGP*, so ist der Anteil der Antworten aus der Ein-Cue-Bedingung (G) höher als der Anteil der korrekten Antworten aus der komplementären Ein-Cue-Bedingung (K). Die 6-jährigen Kinder mit DaE setzen also in der inkongruenten Bedingung stärker auf denjenigen Cue, der sich auch in den Ein-Cue-Bedingungen als intern valider im Vergleich zum komplementären Cue erweist.

Dies kann zum einen auf eine heuristische Verarbeitung zurückgeführt werden, bei der die Auflösungsprozesse primär von einem Default-Cue determiniert werden. Bei einigen Kindern ist der Default-Cue *Genus NGP*, bei anderen der Cue *Konzept*. Es kann aber auch das Ergebnis einer koordinierenden Verarbeitung sein, also einer Verarbeitung, in der beide Cues in die Sprachverarbeitung eingehen und während der Prozessierung bewertet und abgewogen werden. Derjenige Cue, der einen höheren Grad an interner Validität aufweist, setzt sich sozusagen gegenüber dem komplementären Cue durch und bestimmt, an welchen der potenziellen Antezedenten das anaphorische Pronomen schlussendlich angebunden wird. Bei einigen der untersuchten DaE-Kinder wäre es der Cue *Genus NGP*, der eine höhere Validität aufweist, bei anderen der Cue *Konzept*.

Bei den 7-Jährigen kann nicht zwingend auf eine solche positive Korrelation zwischen der inkongruenten Zwei-Cue-Bedingung und den Ein-Cue-Bedingungen geschlossen werden: Die Anteile der korrekten Antworten sind in den

beiden Ein-Cue-Bedingungen gleich hoch und erreichen das Maximum (der Anteil der korrekten Antworten beträgt nahezu 100 %). Die unterschiedlichen Cue-Präferenzen in der inkongruenten Zwei-Cue-Bedingung könnten aber auch bei den 7-Jährigen darauf hindeuten, dass einer der beiden Cues als intern valider gilt – die höhere interne Validität schlägt sich in diesem Alter jedoch nicht mehr in allen Bedingungen nieder, sondern gewinnt insbesondere dann an Bedeutung, wenn ein weiterer, konkurrierender Auflösungscue vorliegt.

Nach den bisherigen Ausführungen lassen sich die wichtigsten Erkenntnisse für die Gruppe der Kinder mit DaE wie folgt zusammenfassen:

- Bereits mit 6 Jahren nutzen Kinder mit DaE die Cues *Genus NGP* und *Konzept* zuverlässig zur Auflösung anaphorischer Personalpronomina.
- Diese Fähigkeit entwickelt sich zwischen dem 6. und 7. Lebensjahr und erreicht in unserem Experiment einen Grad vollständig zielsprachlicher Auflösung.
- Zwischen dem 6. und dem 7. Lebensjahr steigt die interne Validität der untersuchten Cues *Genus NGP* und *Konzept*.
- Die interne Cue-Validität der untersuchten Cues unterliegt bei den DaE-Kindern im untersuchten Altersspektrum einer individuellen Varianz: Bei einigen Kindern ist der Cue *Konzept* der validere, bei anderen der Cue *Genus NGP*.
- Neben der internen Cue-Validität steigt zwischen dem 6. und dem 7. Lebensjahr möglicherweise auch die Automtizität der Cue-Prozessierung und der Cue-Koordination während der Anaphernauflösung.

## Pronomenauflösung DaZ

Die Anteile der korrekten Antworten (in Prozent) für die Gruppe der DaZ-Kinder aufgeteilt nach Bedingung und Altersgruppe sind in Tabelle 3 festgehalten.

Für die Gruppe der DaZ-Kinder gilt, dass sich die Leistungen von denen aus der DaE-Gruppe deutlich unterscheiden. Bei den eindeutigen Items (Typ G, K, und GK) gelingt Kindern mit DaZ die zielsprachliche Auflösung lediglich zu 70,8 %. Betrachtet man alle Antworten, so lassen sich nur 65,5 % davon als zielsprachlich einstufen.

Was die Effekte der Bedingung anbelangt, so konnten zwischen den zwei Ein-Cue-Bedingungen (G und K) keine Unterschiede festgestellt werden, weder bei der Betrachtung der DaZ-Gesamtgruppe noch bei der Betrachtung einzelner Altersgruppen (s. Tabelle 3). Für die Zwei-Cue-Bedingungen (GK und GxK) fallen die Unterschiede in den Leistungen dagegen deutlich aus, sowohl für alle Kinder als auch für die einzelnen Altersgruppen. Die Leistungen in der inkongruenten Zwei-Cue-Bedingung sind dabei wesentlich niedriger

Tabelle 3. Anteile korrekter Antworten in Abhängigkeit von Alter und Bedingung bei Kindern mit DaZ (in %)

	5-Jährige	6-Jährige	7-Jährige	alle
G	65	70	68,8	67,9
K	70	75	68,8	71,4
GK	80	60	81,3	73,2
G x K	45	25	43,8	37,5
G, K, GK	71,7	68,3	72,9	70,8
alle	65	57,5	65,6	65,5

als in der kongruenten Zwei-Cue-Bedingung (s. Tabelle 3). Letztere bewegen sich in etwa auf dem Niveau der Leistungen aus den Ein-Cue-Bedingungen.

Was die Effekte des Alters anbelangt, so herrscht in der Gruppe der DaZ-Kinder folgendes Bild vor: Für die Ein-Cue-Bedingungen konnte zwischen dem 5. und 7. Lebensjahr lediglich eine minimale Entwicklung bei der Prozessorientierung der Auflösungs cues festgestellt werden. Diese Entwicklung zeichnet sich durch einen moderaten Anstieg der Anteile der richtigen Antworten zwischen dem 5. und 6. Lebensjahr und durch einen ebenso moderaten Abfall der Anteile zwischen dem 6. und dem 7. Lebensjahr aus. Verbindet man die einzelnen Werte zu einer Linie, entsteht hier eine leichte Hügelform (s. Abbildung 4). Stärkere Effekte des Alters brachte die Analyse der Items aus den Zwei-Cue-Bedingungen hervor. Die einzelnen Werte lassen sich hier zu einer U-Form verbinden: Die Anteile korrekter Antworten fallen zwischen dem 5. und 6. Lebensjahr ab und steigen zwischen dem 6. und 7. Lebensjahr erneut etwa auf das Niveau der 5-Jährigen an (s. Abbildung 4).

Die Betrachtung der Antworten zu den eindeutigen Items (G, K, GK) brachte eine weitere interessante Beobachtung hervor: Die Anteile der korrekten Antworten aus der Zwei-Cue-Bedingung (GK) sind in der Gruppe der 5- und 7-Jährigen höher als in den Ein-Cue-Bedingungen (G und K), bei den 6-Jährigen ist dies genau umgekehrt (s. Abbildung 4).

Die bisherigen Ausführungen machen deutlich, dass die 6-jährigen Kinder mit DaZ in unserer Studie einen besonderen Stellenwert einnehmen. Zum einen zeigt sich, dass die 6-Jährigen anders als die 5- und 7-Jährigen, anaphorische Pronomina in Texten mit mehreren Cues selbst in der kongruenten Bedingung weniger erfolgreich auflösen als in Texten mit nur einem Cue. Das bedeutet, dass die 5- und 7-Jährigen vom Vorliegen mehrerer Cues profitieren, während multiple Cues selbst bei gleicher Anbindungsrichtung für die 6-Jährigen eine größere Hürde darstellen.

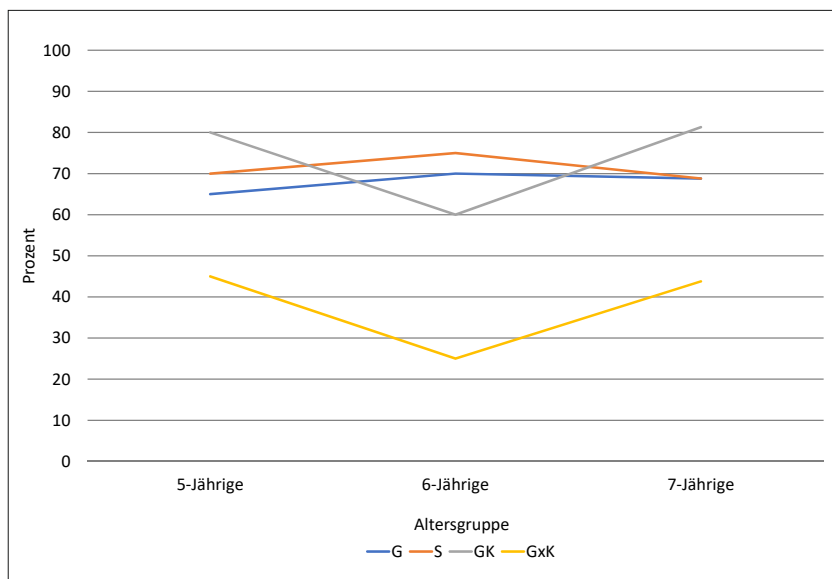


Abbildung 4. Anteile korrekter Antworten in allen Bedingungen (G), (K), (GK) und (GxK) in Abhängigkeit vom Alter bei Kindern mit DaZ (in %)

Folgender Erklärungsansatz ist für diesen Befund vorstellbar. Es ist möglich, dass sich die Cues (hier *Genus NGP* und *Konzept*) bei gleichzeitigem Vorliegen in der Prozessierung durch die 5- und 7-Jährigen gegenseitig verstärken, sodass sich ihre interne Validität erhöht. Sichtbar wird eine solche Verstärkung im Anstieg der korrekten Antworten. In der Ein-Cue-Bedingung ist eine solche stützende Wirkung nicht möglich. Eine derartige Interaktion multipler Auflösungs cues kann als Hinweis auf die Fähigkeit der 5- und 7-Jährigen gelten, mehrere Cues während der Auflösung anaphorischer Personalpronomina zu koordinieren – ohne Koordination wäre eine gegenseitige Verstärkung der Cue-Validität nicht möglich. Diese Befunde unterscheiden sich leicht von den Ergebnissen bisheriger Studien zur Cue-Koordination im DaZ-Erwerb. So fanden zum Beispiel Klages & Gerwien (2015) heraus, dass im DaZ-Erwerb erst ab dem 6. Lebensjahr von der Koordination multipler Hinweise ausgegangen werden kann.<sup>6</sup>

6 Klages & Gerwien (2015) untersuchten in ihrer Studie die Koordination von dem Cue *Genus NGP* und der syntaktischen Rolle der vorausgegangenen Ausdrücke. Ob die Unterschiede zwischen dieser Studie und der vorliegenden Studie den implementierten Untersuchungsmethoden (bei Klages & Gerwien die Visual World Methode, hier eine Entscheidungsaufgabe), dem Untersuchungsgegenstand (bei Klages & Gerwien die Koordination von *Genus NGP* und *Position im Satz* bzw. *syntaktische Rolle des Antezedenten*, hier *Genus NGP* und *Konzept*) oder der hier untersuchten

In der vorliegenden Studie scheinen die 6-jährigen Kinder aus unserer Stichprobe in Bezug auf die Cue-Koordination hingegen vor einer größeren Herausforderung zu stehen als die jüngeren oder älteren Kinder. Das könnte daran liegen, dass diese Altersstufe sich in einem Erwerbsstadium befindet, in dem etwas anderes wie beispielsweise die Prozessierung eines einzelnen Cues im Vordergrund steht. Dafür sprächen auch die leicht erhöhten Anteile der korrekten Antworten in den Ein-Cue-Bedingungen (G und K) in dieser Altersgruppe. Sind die Kinder also erwerbsbedingt gerade dabei, die Prozessierung einzelner Cues auszubauen, zu stabilisieren oder zu automatisieren, so benötigen diese Prozesse einen Großteil der kindlichen Aufmerksamkeit und der kognitiven Ressourcen. Texte mit mehreren Cues können sich in diesem Erwerbsstadium negativ auf die Cue-Verarbeitung auswirken, zum Beispiel weil die Anwesenheit eines weiteren Cues zusätzliche Ressourcen beansprucht. Dies bedeutet also, dass auch bei den 6-Jährigen mehrere Cues in die Verarbeitungsprozesse eingehen, dies sich aber erwerbsbedingt negativ auf deren Verarbeitung auswirkt.

Kommen wir nun zur Diskussion der Werte der 5- und 7-jährigen DaZ-Kinder. In diesen Altersgruppen sind die Anteile der Antworten in jeder Bedingung vergleichbar. Heißt das etwa, dass auch die Verarbeitungsprozesse in diesen beiden Altersgruppen vergleichbar sind? Die beobachtete U-Form in den Werten in einem bestimmten Altersspektrum (hier zwischen dem 5. bis 7. Lebensjahr) wird in der Spracherwerbsforschung als Hinweis auf unterschiedliche Verarbeitungsprozesse in den Altersgruppen am jeweiligen Ende des Alterspektrums (hier 5 und 7 Jahre) gedeutet (z. B. Karmiloff-Smith, 1992).

Eine Antwort auf die Frage, wie die Verarbeitungsprozesse in der kongruenten Zwei-Cue-Bedingung in der Gruppe der 5- und der 7-Jährigen aussehen, könnte die Betrachtung der individuellen Leistungen der untersuchten Kinder liefern (s. Abbildung 5).

Ein erster Blick auf die Leistungen der einzelnen Kinder mit DaZ zeigt, dass Kinder mit DaZ sich insgesamt heterogener als Kinder mit DaE verhalten – das gilt sowohl für die Gesamtgruppe als auch für die Kinder innerhalb einer Altersgruppe: In jeder Altersgruppe mit DaZ gibt es Kinder, die überzufällig korrekt antworten und auch Kinder, die sich zufällig verhalten. Das bedeutet, dass die DaZ-Gruppe (im Vergleich zu der DaE-Gruppe) insgesamt weniger generalisierbare Aussagen zulässt. Eine Interpretation der Daten werde ich trotzdem vornehmen, betrachte sie jedoch als Hypothesenformulierung, die es künftig zu überprüfen gilt.

Was die Leistungen der 5- und der 7-jährigen Kinder in der kongruenten Zwei-Cue-Bedingung anbelangt, so offenbart die Betrachtung der individuellen

---

kleinen Stichprobe geschuldet sind, kann nicht beantwortet werden. Erst zukünftige Forschung würde zur Klärung beitragen.

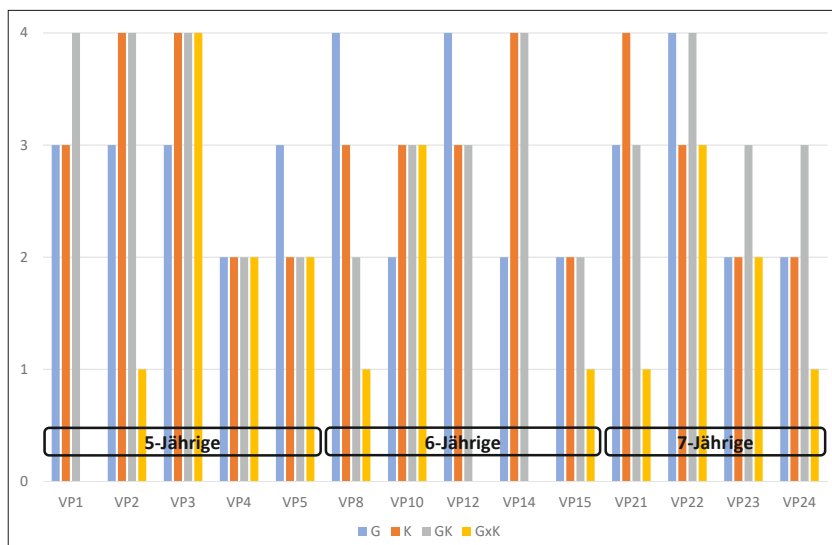


Abbildung 5: Korrekte Antworten für jedes Kind mit DaZ (absolute Zahlen)

Leistungen in den beiden Altersgruppen zwei Hauptkenntnisse: Es konnte zum einen festgestellt werden, dass die Performanzmuster der 5-Jährigen sich von den Performanzmustern der 7-Jährigen unterscheiden: Die jüngsten Kinder mit DaZ lösen Pronomina in der kongruenten Zwei-Cue-Bedingung nur dann überzufällig korrekt auf (mindestens drei von vier Antworten), wenn sie auch in beiden Ein-Cue-Bedingungen die Pronomina auf der Grundlage der vorliegenden Cues auflösen (VP1, VP2, VP3). Zwar weisen zwei der 7-Jährigen (VP21, VP22) dieses Verarbeitungsmuster auf, in dieser Altersgruppe sind aber auch Kinder vertreten, die sich überzufällig korrekt in der kongruenten Zwei-Cue-Bedingung verhalten, ohne in den Ein-Cue-Bedingungen die vorliegenden semantischen Cues systematisch zu berücksichtigen (VP23, VP24). Dies kann als ein weiterer Hinweis auf das Vorhandensein von jeweils unterschiedlichen Verarbeitungsstrategien in den zwei Altersgruppen gewertet werden. Auf die Unterschiede kommen wir später dezidiert zurück.

Zum anderen wird deutlich, dass drei von den 5-Jährigen in der kongruenten Zwei-Cue-Bedingung zu 100 % korrekt antworten und zwei nach dem Zufallsprinzip vorgehen, während die vier 7-jährige Kinder zu mindestens 75 % korrekt antworten. Die vergleichbaren Gesamtwerte zwischen den zwei Altersgruppen sind also auf unterschiedliche Verteilungen der korrekten Antworten innerhalb der Altersgruppen zurückzuführen.

Kommen wir nun zurück zur Diskussion der Verarbeitungsstrategien bei den 5- und 7-jährigen Kindern in der kongruenten zwei-Cue-Bedingung. Oben habe ich argumentiert, dass Kinder beider Altersgruppen die zwei vorliegenden Cues *Genus NGP* und *Konzept* koordinieren, in der Koordination

eine gegenseitige Stützung der beiden Cues stattfindet und zur Erhöhung der internen Validität der Cues führt. Eine solche Koordination wird bei den 5-Jährigen offenbar jedoch nur dann implementiert, wenn der Grad der Prozessierung jedes einzelnen Cues (in der Ein-Cue-Bedingung) einen gewissen Wert erreicht, also einen bestimmten Grad interner Validität besitzt<sup>7</sup>. Bei den 7-Jährigen scheint eine solche Koordination bereits bei einem niedrigeren Grad interner Validität der Cues möglich zu sein, und zwar bereits auch dann, wenn die Kinder in den Ein-Cue-Bedingungen zufällig performieren. 5-jährige Kinder, die in den Ein-Cue-Bedingungen zufällig performieren, profitieren dagegen noch nicht von dem Vorhandensein zweier Cues, wie die Ergebnisse von VP 4 und VP5 nahelegen.

Zusammenfassend können die vorliegenden Befunde für die DaZ-Gruppe also folgendermaßen interpretiert werden: Die jüngsten Kinder mit DaZ aus unserer Stichprobe implementieren Cue-basierte Pronomenauflösung und können währenddessen auch multiple Cues koordinieren, die Koordination wird jedoch erst dann initiiert, wenn ein bestimmter Grad an interner Cue-Validität erreicht ist. Individuen, bei denen die untersuchten Cues noch eine zu niedrige interne Validität besitzen, initiieren keine oder nur wenige koordinierende Prozesse. Stattdessen stützen sie die Pronomenauflösung auf einzelne Cues. Diese Cues könnten als eine Art Default betrachtet werden, die hierbei ablaufenden Prozesse wären dann als heuristisch zu betrachten. Mit steigendem Kontakt zur Zielsprache (hier unsere 6-Jährigen) werden die koordinierenden Verarbeitungsprozesse reduziert und die Kinder konzentrieren sich stärker auf den inputbasierten Ausbau der internen Validität der einzelnen Cues, verlassen also die Default-Strategie. Nach diesem Erwerbsstadium werden wieder stärker die koordinierenden Prozesse aufgenommen (in unserer Studie bei den 7-Jährigen), selbst dann, wenn die interne Validität der Auflösungshinweise einen niedrigen Grad aufweist.

Die soeben beschriebene Entwicklung (von einer stärker heuristisch-koordinierenden über eine auf einem Cue basierenden zu einer inputgestützten koordinierenden Verarbeitung) korreliert bei den Kindern mit DaZ nicht absolut mit einem bestimmten Alter. Wie bereits oben erwähnt, verhalten sich die Kinder mit DaZ sehr heterogen – insgesamt und auch innerhalb der Altersgruppen. Die Entwicklung ist in der beobachteten Altersspanne (5 bis 7 Jahre) anzusiedeln, wobei die heuristisch-koordinierenden Prozesse eher von den jüngeren Kindern und die inputgestützten koordinierenden eher von den älteren Kindern implementiert werden. Aber auch bei den 7-Jährigen konnten noch Kinder gefunden werden, die die interne Validität der einzelnen Cues ausbauen (VP 21).

---

7 Dieser Wert führt in der vorliegenden Studie zu mindestens 75 % korrekten Antworten in jeder Ein-Cue-Bedingung.



Bei der Betrachtung der individuellen Leistungen der DaZ-Kinder fällt weiterhin auf, dass in jeder Altersgruppe Kinder sind, die in den Ein-Cue-Bedingungen zufällig antworten, das heißt, bei der Pronomenauflösung weder den Cue *Genus NGP* noch den Cue *Konzept* systematisch berücksichtigen. Solche Kinder konnten in der DaE-Gruppe nicht gefunden werden. Gehen diese Kinder etwa gänzlich unsystematisch bei der Pronomenauflösung vor?

Eine Postanalyse zeigte, dass die Kinder keineswegs unsystematisch vorgehen. Wie in Abbildung 6 ersichtlich ist, lösen diese Kinder die anaphorischen Personalpronomina sehr wohl unter Beachtung von sprachlichen Auflösungs-cues auf, nämlich der Cues *syntaktischen Rolle / Nennungsposition* der vorausgehenden Ausdrücke. Diese Cues steuern hier die Präferenz, mit der die rezipierten Pronomina angebunden werden.

Die Anbindungspräferenzen lassen sich wie folgt beschreiben: Während bei dem 5-jährigen DaZ-Kind die Anbindung an das vorausgegangene Subjekt (das hier mit der Erstnennung konfundiert ist) dominiert, bindet das Kind im Alter von 6 Jahren die anaphorischen Personalpronomina an das vorausgegangene Objekt an. Die zwei 7-Jährigen präferieren erneut eine Subjektanbindung. In Anlehnung an die obige Ausführung zur u-förmigen Sprachentwicklung – auch hier zeigt der Wechsel der Anbindungspräferenz bei den 6-Jährigen eine solche Entwicklung an – gehe ich davon aus, dass die Anbindung an die vorausgegangenen Subjekte bei den 5- und den 7-Jährigen nicht auf gleichartige Verarbeitungsprozesse zurückzuführen ist, und interpretiere die Ergebnisse wie folgt: Das jüngere Kind implementiert bei der Pronomenauflösung eine heuristische Verarbeitung, wendet also eine sog. Default-Strategie an. Der Default-Wert stellt dabei das vorausgegangene Subjekt bzw. die Erstnennung dar. Diese Strategie ist möglicherweise noch nicht (oder nicht ausschließlich) inputbasiert, sondern in der kindlichen Präferenz, Erstnennungen als prominent zu betrachten (z. B. Song & Fisher, 2005, 2007) verankert. Ich nenne sie die „distributionelle Strategie“. In einem späteren Stadium (hier das 6-jährige Kind) wird diese Default-Strategie ‚überprüft‘, indem nun verstärkt das vorausgegangene Objekt / Zweitnennung in den Fokus der Aufmerksamkeit genommen wird. Möglicherweise ist die Nennungsnähe (Recency) zwischen Pronomen und dem Objekt / Zweitgenannten für den Verlust der Prominenz der Erstnennungen und somit für den Aufbruch der früheren Default-Strategie mitverantwortlich. Während dieser Entwicklungsphase (d. h. der Überprüfungsphase) lenken die Kinder ihre Aufmerksamkeit stärker auf den sie umgebenden Input, um Hinweise für eine inputbasierte Anbindung zu finden. Während dieser Phase nehmen sie stärker als zuvor die Objekte als potenzielle Antezedenten in den Blick. Nach Beendigung dieser ‚Überprüfungsphase‘, in der sie festgestellt haben, dass Objekte / Nicht-Erstnennungen auch den Antezedenten eines anaphorischen Personalpronomens darstellen können, allerdings weniger häufig als Subjekte / Erstnennungen, implementieren die Kinder eine neue,

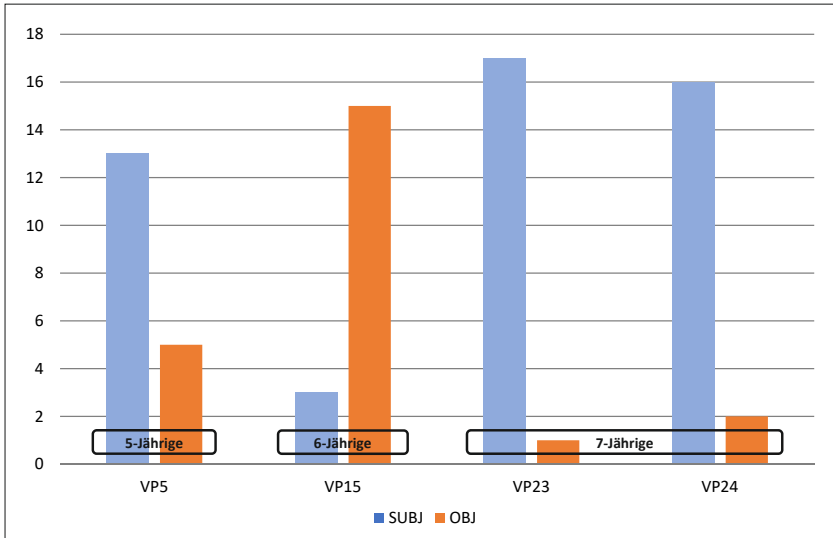


Abbildung 6: Anbindungspräferenzen für DaZ-Kinder, die in den Ein-Cue-Bedingung die semantischen Cues nicht systematisch berücksichtigen (absolute Zahlen für alle Items)

inputbasierte Verarbeitungsstrategie, nach der Personalpronomina im Deutschen bevorzugt auf vorausgegangene Subjekte / Erstnennungen Bezug nehmen. Ich nenne diese Strategie die „syntaktische Strategie“.

Vor dem Hintergrund der bisherigen Ausführungen lässt sich mit Blick auf den Erwerb der Auflösung anaphorischer Personalpronomina durch DaZ-Kinder folgende Entwicklungssequenz formulieren:

1. Distributionelle Default-Strategie: anaphorische Personalpronomina werden an vorausgegangene Erstnennungen angebinden.
2. Überprüfung des Defaults: anaphorische Personalpronomina werden an die komplementären Antezedenten, d. h. an die in der unmittelbaren Nähe des Pronomens stehenden angebinden.
3. Inputbasierte Syntax-Strategie: anaphorische Personalpronomina werden an vorausgegangene Subjekte / Erstnennungen angebinden.
4. Sensibilisierung für die Cues *Genus NGP* und *Konzept* und Koordination dieser Cues, wenn ein bestimmter Grad der internen Cue-Validität erreicht ist.
5. Inputgestützte Überprüfung und Ausbau der internen Validität der Cues *Genus NGP* und *Konzept* und deren Verarbeitung insbesondere in Ein-Cue-Kontexten.
6. Ausbau und Stabilisierung koordinierender Verarbeitungsprozesse auch bei einem niedrigen Grad der internen Validität der Cues *Genus NGP* und *Konzept*.

Nun sollen die wichtigsten Ergebnisse für die Kinder mit DaZ zusammengefasst werden:

- Bereits mit 5 Jahren können Kinder mit DaZ die Cues *Genus NGP* und *Konzept* zur Auflösung anaphorischer Personalpronomina in Texten nutzen. Diese Fähigkeit ist jedoch von individuellen Erwerbsbedingungen abhängig und auch im Alter von 7 Jahren noch nicht zwingend bei jedem Kind mit DaZ zu erwarten.
- Der Nutzung der Cues *Genus NGP* und *Konzept* geht im DaZ-Kontext die Berücksichtigung distributioneller und syntaktischer Auflösungsclues voraus.
- Die Entwicklung der Fähigkeiten zur Pronomenauflösung ist bei Kindern mit DaZ vermutlich stark von den individuellen Erwerbsbedingungen abhängig, was in einem weitaus heterogeneren Gesamtbild im Vergleich zu den DaE-Kindern resultiert. Hierzu zählen zum Beispiel die Quantität und Qualität des sprachlichen Inputs (Kaltenbacher & Klages, 2006; Pagonis, 2022; Czinglar et al., 2017), die Erfahrung mit Texten und den dort enthaltenen Kohäsionsmitteln (Bryant & Noschka, 2015) oder die Erstsprache der Kinder (Kaltenbacher, 2015).

## 5 Zusammenfassung und didaktische Implikationen

In dem vorliegenden Beitrag wurden die Befunde einer explorativen Studie präsentiert und daraus einige Hypothesen zum Erwerb der Fähigkeiten zur Pronomenauflösung bei 5- bis 7-jährigen Kindern mit DaZ und bei 6- bis 7-jährigen Kindern mit DaE abgeleitet. Zu den Hauptergebnissen zählt zum einen, dass der Erwerbskontext (DaE vs. DaZ) einen Einfluss auf den Erwerb der Fähigkeiten zur Pronomenauflösung hat: Im DaZ-Kontext erfolgt der Erwerb heterogener im Vergleich zum Erwerb im DaE-Kontext, was vermutlich den heterogenen Erwerbsbedingungen im DaZ-Kontext geschuldet ist. Zum anderen wurde deutlich, dass das Maß der korrekten Lösungen im DaZ-Kontext niedriger ist als im DaE-Kontext, was auf die Sinnhaftigkeit einer gezielten sprachdidaktischen Intervention für Kinder mit DaZ hindeutet. Des Weiteren wurde sichtbar, dass Kinder mit DaZ auf Auflösungsstrategien zurückgreifen, die bei den untersuchten Kindern mit DaE nicht (mehr) beobachtet wurden. Es handelt sich um Strategien, bei denen insbesondere die distributionellen und syntaktischen Auflösungsclues wie Position oder syntaktische Rolle der Antezedenten im vorausgegangenen Satz eine Rolle spielen. Diese Strategien gehen laut unseren Befunden im DaZ-Erwerb der Verarbeitung des Cues *Genus-NGP* und des Cues *Konzept* voraus und zeigen also frühere Erwerbsstadien an. Dass im DaZ-Erwerb die Position / syntaktische Rolle des

Antezedenten den semantischen oder konzeptuellen Strategien vorausgeht, ist keineswegs selbstverständlich. Zwar gehen einige Forscherinnen von einer sehr frühen Implementierung dieser Strategie im kindlichen Spracherwerb aus (Song & Fisher, 2005, 2007 zu Englisch), andere nehmen dagegen an, dass semantische Cues im Spracherwerb früher erworben werden als andere Cues (Arnold et al., 2005). Unserer Studie zufolge scheint diese Strategie im DaZ-Erwerb eine erwerbsfrühe Strategie zu sein. Inwiefern diese Strategie auch im DaE-Erwerb den semantischen und konzeptuellen Strategien vorausgeht, kann aufgrund der vorliegenden Studie nicht geschlussfolgert werden. Hierfür wäre weitere Forschung notwendig. Frühere Studien zum L1-Erwerb legen allerdings den Schluss nahe, dass auch im L1-Erwerb bereits sehr junge Kinder bei der Informationsverarbeitung auf die Nennungsabfolge der Informationen sensibel reagieren (zu L1-Deutsch Dimroth & Narasimhan, 2012) und anaphorische Pronomina bevorzugt an Erstnennungen anbinden (zu L1-Englisch Song & Fisher, 2005 und 2007, zu L1-Bulgarisch und Deutsch Bittner & Kühnast, 2012).

Interessanterweise zeigt die vorliegende Studie, dass die Mehrheit der Kinder sowohl im DaZ- als auch im DaE-Erwerb bei der Pronomenauflösung systematisch vorgeht, d. h. irgendwelche Cues berücksichtigt und (in späteren Erwerbstadien) sogar koordiniert. Das bedeutet, dass Kinder im Erwerb nach sprachlichen Regularitäten suchen und sich auf diese während der Sprachverarbeitung stützen. Für eine sprachdidaktische DaZ-Intervention, deren Sinnhaftigkeit ich in den heterogenen und weniger häufig zielsprachlichen Reaktionen der DaZ-Kinder begründet sehe, bedeutet das, dass dort sprachliche Regularitäten gezielt berücksichtigt werden müssen. In der Sprachförderung sollte also den Kindern ein Input angeboten werden, in dem diese Regularitäten für die Kinder erkennbar sind und erworben werden können (zum Beispiel eine in den Usage-based-Ansätzen verankerte Sprachförderung). Sollte während der Sprachintervention auch der natürliche Erwerbsverlauf berücksichtigt werden, so sollte die Sprachförderung zunächst den Erwerb der distributionellen / syntaktischen Cues unterstützen und den Erwerb des Cues *Genus-NGP* bzw. des Cue *Konzept* erst später in den Fokus nehmen.

## Literatur

- Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S. & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, 76(1), 13–26.
- Arnold, J. E., Brown-Schmidt, S., Trueswell, J. C. & Fagnano, M. (2005). Children's use of gender and order-of-mention during pronoun comprehension. In J. C. Trueswell & M. K. Tanenhaus (Hrsg.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (S. 261–281). MIT Press.
- Arnold, J. E., Brown-Schmidt, S. & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, 22(4), 527–565.
- Bittner, D. (2007). Influence of animacy and grammatical role on production and comprehension of intersentential pronouns in German L1-acquisition. In D. Bittner & N. V. Gagarina (Hrsg.), *Intersentential pronominal reference in child and adult language* (S. 103–138). Zentrum für allgemeine Sprachwissenschaft.
- Bittner, D. & Kuehnast, M. (2012). Comprehension of intersentential pronouns in child German and child Bulgarian. *First Language*, 32(1–2), 176–204.
- Bosch, P. & Umbach, C. (2007): Reference determination for demonstrative pronouns. In D. Bittner & N. V. Gagarina (Hrsg.), *Intersentential pronominal reference in child and adult language* (S. 39–51). Zentrum für allgemeine Sprachwissenschaft.
- Bryant, D. & Noschka, N. (2015). Personal- und Demonstrativpronomen im Sprachverstehensprozess. Untersuchungen zum Erwerb funktionaler Anapherndistribution bei DaM, DaF und DaZ. In H. Klages & G. Pagonis (Hrsg.), *Linguistisch fundierte Sprachförderung und Sprachdidaktik. Grundlagen, Konzepte, Desiderate* (S. 17–46). De Gruyter.
- Clahsen, H. & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, 27(1), 3–42.
- Clahsen, H. & Felser, C. (2018). Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*, 40(3), 693–706.
- Crawley, R. A., Stevenson, R. J. & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, 19(4), 245–264.
- Czinger, C., Rüdiger, J. O., Korecky-Kröll, K., Uzunkaya-Sharma, K. & Dressler, W. U. (2017). Inputfaktoren im DaZ-Erwerb von sukzessiv bilingualen Kindern mit L1 Türkisch. In I. Fuchs, S. Jeuk & W. Knapp (Hrsg.), *Mehrsprachigkeit: Spracherwerb, Unterrichtsprozesse, Seiteneinstieg* (S. 15–34). Klett Fillibach.

- Dimroth, C. & Narasimhan, B. (2012). The acquisition of information structure. In M. Krifka & R. Musan (Hrsg.), *The expression of information structure* (S. 319–362). De Gruyter.
- Ehrlich, K. (1980). Comprehension of pronouns. *Quarterly Journal of Experimental Psychology*, 32(2), 247–255.
- Ellert, M. (2011). *Ambiguous Pronoun Resolution in L1 and L2 German and Dutch*. Ponsen & Looijen.
- Frederiksen, J. R. (1981). Understanding anaphora: Rules used by readers in assigning pronominal referents. *Discourse Processes*, 4(4), 323–347.
- Gagarina, N. V. (2007). The hare hugs the rabbit. He is white . . . Who is white? Pronominal anaphora in Russian. In D. Bittner & N. V. Gagarina (Hrsg.), *Intersentential pronominal reference in child and adult language* (S. 139–170). Zentrum für allgemeine Sprachwissenschaft.
- Gernsbacher, M. A. & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, 27(6), 699–717.
- Järvikivi, J., Pyykkönen-Klauck, P., Schimke, S., Colonna, S. & Hemforth, B. (2014). Information structure cues for 4-year-olds and adults: Tracking eye movements to visually presented anaphoric referents. *Language, Cognition and Neuroscience*, 29(7), 877–892.
- Järvikivi, J., van Gompel, R. P., Hyönä, J. & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject preference accounts. *Psychological Science*, 16(4), 260–264.
- Kaltenbacher, E. (2015). Einfluss der Erstsprache im frühen kindlichen Zweitspracherwerb. Erwerb des deutschen Genussystems durch russische und türkische Kinder. In H. Rösch & J. Webersik (Hrsg.), *Deutsch als Zweitsprache – Erwerb und Didaktik. Beiträge aus dem 10. Workshop „Kinder mit Migrationshintergrund“* (S. 55–72). Klett Fillibach.
- Kaltenbacher, E. & Klages, H. (2006). Sprachprofil und Sprachförderung bei Vorschulkindern mit Migrationshintergrund. In B. Ahrenholz (Hrsg.), *Zweitspracherwerb und curriculare Dimensionen: empirische Untersuchungen zum Deutschlernen in Kindergarten und Grundschule* (S. 80–97). Stauffenburg.
- Karmiloff-Smith, A. (1992). *Beyond Modularity. A Developmental Perspective on Cognitive Science*. MIT Press.
- Klages, H. & Gerwien, J. (2015). Verstehen anaphorischer Personalpronomina im DaZ- und DaM-Erwerb. In H. Klages & G. Pagonis (Hrsg.), *Linguistisch fundierte Sprachförderung und Sprachdidaktik – Grundlagen, Konzepte, Desiderate* (S. 71–98). De Gruyter.
- Klages, H. & Gerwien, J. (2020). Referential coherence: Children’s understanding of pronoun anaphora. Insights from mono- and bilingual language acquisition. In N. V. Gagarina & R. Musan (Hrsg.) *Referential and relational discourse coherence in adults and children* (S. 105–138). De Gruyter.

- McKoon, G. & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99(3), 440–466.
- Pagonis, G. (2022). Implizite Formfokussierung in der elementarpädagogischen Sprachförderung. In K. Madlener-Charpentier & G. Pagonis (Hrsg.), *Aufmerksamkeitslenkung und Bewusstmachung in der Sprachvermittlung. Kognitive und didaktische Perspektiven auf Deutsch als Erst-, Zweit- und Fremdsprache* (S. 129–162). Narr Francke Attempto.
- Pan, H.-Y., Schimke, S. & Felser, C. (2015). Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism*, 19(3), 298–313.
- Schwarz-Friesel, M. (2000). Indirekte Anaphern in Texten. Studien zur domänengebundenen Kohärenz und Referenz im Deutschen. Niemeyer.
- Shillcock, R. (1982). The on-line resolution of pronominal anaphora. *Language and Speech*, 25(4), 385–401.
- Song, H. & Fisher, C. (2005). Who’s “she”? Discourse prominence influences preschoolers’ comprehension of pronouns. *Journal of Memory and Language*, 52(1), 29–57.
- Song, H. & Fisher, C. (2007). Discourse prominence effects on 2.5-year-old children’s interpretation of pronouns. *Lingua*, 117(11), 1959–1987.
- Sorace, A. & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22(3), 339–368.
- Tanenhaus, M. K. & Trueswell, J. C. (1995). Sentence Comprehension. In J. L. Miller & P. D. Eimas (Hrsg.), *Speech, language, and communication. Handbook of perception and cognition* (S. 217–262). Academic Press.

## Anhang

### Textstimuli: Testitems

#### Typ 1: Ein-Cue Genus NGP (G)

1. Die Leute spielen Verstecken. Die Hexe sucht den Koch. Sie hat neue Schuhe an.
2. Die Leute spielen Verstecken. Der Pirat sucht die Fee. Er hat neue Schuhe an.
3. Die Leute wollen ausgehen. Der Ritter ruft die Königin an. Sie hat ein neues Handy.
4. Die Leute wollen ausgehen. Die Fee ruft den Ritter an. Er hat ein neues Handy.

### Typ 2: Ein-Cue *Konzept* (K)

5. Die Leute wollen ausgehen. Die Königin ruft die Krankenschwester an. Sie wählt die 102030.
6. Die Leute wollen ausgehen. Der Pirat ruft den Koch an. Er wählt die 313233.
7. Die Leute spielen Verstecken. Die Fee sucht die Hexe. Sie versteckt sich hinter einem Vorhang.
8. Die Leute spielen Verstecken. Der Ritter sucht den Koch. Er versteckt sich hinter einem Vorhang.

### Typ 3: Zwei-Cues – kongruent (GK)

9. Die Leute wollen ausgehen. Die Hexe ruft den Pirat an. Sie wählt die 987654.
10. Die Leute wollen ausgehen. Der Ritter ruft die Hexe an. Er wählt die 432109.
11. Die Leute spielen Verstecken. Der Koch sucht die Königin. Sie versteckt sich hinter einem Vorhang.
12. Die Leute spielen Verstecken. Die Fee sucht den Ritter. Er versteckt sich hinter einem Vorhang.

### Typ 4: Zwei-Cues – inkongruent (GxK)

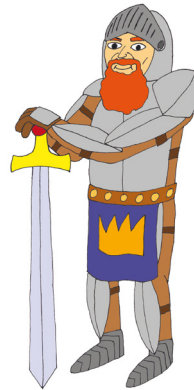
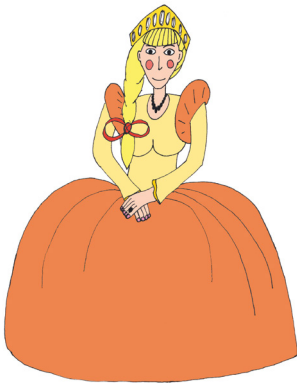
13. Die Leute wollen ausgehen. Die Fee ruft den Pirat an. Sie geht nicht dran.
14. Die Leute wollen ausgehen. Der Koch ruft die Fee an. Er geht nicht dran.
15. Die Leute spielen Verstecken. Der Ritter sucht die Königin. Sie zählt zuerst bis zehn.
16. Die Leute spielen Verstecken. Die Hexe sucht den Ritter. Er zählt zuerst bis zehn.

### Textstimuli: Kontrollitems

1. Die Leute spielen Verstecken. Die Königin sucht die Hexe. Sie hat braune Haare.
2. Die Leute spielen Verstecken. Der Koch sucht den Pirat. Er hat braune Haare.



Beispiele visueller Stimuli





---

Rolf Koepfel

# Unbeabsichtigte Verursachung: Zu Widerständen beim Erwerb einer Konstruktion in der Fremdsprache Deutsch und Möglichkeiten ihrer Vermittlung

---

**Zusammenfassung** Konstruktionen Unbeabsichtigter Verursachung sind Gegenstand der Typologie, ihr Ausdruck im Deutschen durch ergative Verben mit Dativ wird in einschlägigen wissenschaftlichen Grammatiken intensiv diskutiert, doch in der Didaktik des Deutschen als Fremd- und Zweitsprache findet das Phänomen keine Berücksichtigung. Dabei zeigen Daten fortgeschrittener Deutschlerner, dass die Konstruktion allein durch implizites Lernen aus dem Input schlecht erworben wird, mit Folgen für die Rezeption und mögliche Konzeptualisierung so ausgedrückter Ereignisse. In der Produktion fehlt entsprechend ein ökonomisches Ausdrucksmittel. Der Artikel geht den Gründen für die Erwerbsschwierigkeiten nach und macht einen konstruktionsgrammatisch inspirierten Vorschlag, wie Inputgestaltung und explizitere Verfahren der Bewusstmachung den Erwerb dieser und der eng verwandten Konstruktion Nichtverhinderter Ereignisse erleichtern können.

**Schlüsselwörter** Konstruktionsgrammatik; Fremdsprachenunterricht; ergative / unakkusative Verben; Dativ; Agentivität; explizites Lernen; implizites Lernen

## 0 Einleitung

Unbeabsichtigte Verursachung wie in *Mir ist das Handy heruntergefallen* wird in der Typologie untersucht (Kittilä, 2005), in der kognitiven Linguistik angesprochen (u. a. Fausey & Boroditsky, 2011), wissenschaftliche Grammatiken des Deutschen diskutieren die Argumentstruktur von Verben, die im Deutschen zu ihrem Ausdruck verwendet werden (u. a. Eisenberg, 2020),

besonders die zentrale Rolle des Dativs, doch in der Didaktik des Deutschen als Fremd- und Zweitsprache findet das Phänomen keine Berücksichtigung. Das ist ein Mangel, denn die Konstruktion setzt dem impliziten Lernen allein aus dem Input beträchtliche Widerstände entgegen. In der Rezeption bietet sich Lernenden in Analogie zu anderen Verwendungsweisen des Dativs eine unterspezifizierte Interpretation an mit Folgen für die Kognition solcher Ereignisse, während in der Produktion ein ökonomisches Ausdrucksmittel für den Ereignistyp fehlt. Deshalb wird hier ein Vorschlag zur Vermittlung unterbreitet, der Möglichkeiten von Inputgestaltung und Bewusstmachung vor dem Hintergrund einer Analyse der Lernprobleme vorstellt.

Abschnitt 1 setzt die typologische Beschreibung durch Kittilä (2005) als linguistischen Ausgangspunkt. Kittilä stellt prototypische Eigenschaften der Konstruktion heraus, die auch der didaktischen Vermittlung einen geeigneten Zugang eröffnen. Im Anschluss werden der spezifische Ausdruck Unbeabsichtigter Verursachung im Deutschen typologisch verortet und die eng verwandte Konstruktion Nichtverhinderter Ereignisse einbezogen. Der Analyse liegt ein konstruktionsgrammatischer Ansatz zugrunde. Abschnitt 2 weist anhand von Daten weit fortgeschrittener L2-Sprecher die Schwierigkeiten des Erwerbs beider Konstruktionen nach und versucht ihre Ursachen zu ergründen, wobei auch subjektive Lerneräußerungen herangezogen werden. Darauf aufbauend wird in Abschnitt 3 der Vermittlungsvorschlag vorgestellt, der sich ebenfalls im Rahmen konstruktionsgrammatischer Überlegungen bewegt.

## 1 Zur Konstruktion der Unbeabsichtigten Verursachung

Mit (1) liegt eine Konstruktion vor, die Kittilä (2005) in seiner typologischen Untersuchung in Übernahme eines Terminus aus Haspelmaths (1993) Grammatik des Lesgischen *Involuntary Agent Construction* nennt; im Folgenden wird von Konstruktion der Unbeabsichtigten Verursachung gesprochen.

(1) Dem Besucher ist die kostbare Vase zerbrochen.

### 1.1 Zentrale Eigenschaften der Konstruktion

1. Die Konstruktion kodiert die unerwartet niedrige Agentivität an sich hoch agensfähiger Entitäten. Deshalb werden kanonische Agenten bevorzugt, wie die abnehmende Akzeptabilität in (2) zeigt; wenig überraschend sind verursachende Kräfte wie Winde usw. gänzlich ausgeschlossen.

(2) dem Mann / ?dem Baby / ??der Katze ist der Teller zerbrochen  
(Kittilä, 2005, S. 409)

2. Ein entsprechender transitiver Handlungssatz wie *Der Mann hat den Teller zerbrochen* ist möglich, doch im Gegensatz zu letzterem wird 3. ausgedrückt, dass der Dativreferent nicht mit Absicht gehandelt hat und keine Kontrolle über das Ereignis hatte.

Im Deutschen wird die agensfähige Entität im Dativ kodiert. Zieht man den Proto-Rollen-Ansatz von Dowty (1991) heran, kommen dieser Dativrolle drei Agens-Eigenschaften zu: (i) Sentience / Perception (die Fähigkeit, das Ereignis wahrzunehmen), (ii) Verursachung sowie (iii) Bewegung hinsichtlich eines anderen Partizipanten. Nur eine Proto-Agens-Eigenschaft fehlt, nämlich Volition bzw. Handlungskontrolle<sup>1</sup>. Sofern Bewegungsverben vorliegen, ist nach Takahashi (2017) eine typische Bedingung für die Interpretation als unbeabsichtigte Verursachung, dass die Bewegung von der Dativperson aus gesehen zentrifugal verläuft wie in (3a); bei zentripetaler Bewegung drückt der Dativ lediglich Affiziertheit im Sinne des *Dativus commodi / incommodi* (Dativ des Nutznießers / Geschädigten) aus:

- (3a) Dabei war ihm die laufende Säge aus der linken Hand gerutscht ... (Hannoversche Allgemeine vom 20.7.2009, zitiert nach Takahashi, 2017, S. 368)
- (3b) ... und das Wasser den Leuten in den Keller gelaufen ist (Braunschweiger Zeitung vom 2.6.2008, zitiert nach Takahashi, 2017, S. 369)

4. Der Verursacher ist vom Resultat des Ereignisses negativ betroffen. Das ist für Kittilä (2005, S. 414) dadurch motiviert, dass Menschen Ereignisse in der Regel nur dann willentlich auslösen, wenn sie für sie vorteilhaft sind; bei unbeabsichtigter Auslösung ist ein nachteiliger Effekt eher wahrscheinlich.

Unbeabsichtigte Verursachung nimmt im Deutschen von ihrem Ausdruck her eine Mittelstellung zwischen dem Ausdruck typischer transitiver Handlungen und typischer intransitiver Zustandsveränderungs-Ereignisse ein, wie Tabelle 1 illustriert. Die Zweistelligkeit ist eine an Transitivität erinnernde Eigenschaft ebenso wie die bevorzugte Abfolge des agentiveren Arguments vor dem Argument mit Patienceigenschaften, das Verb aber ist intransitiv und bildet das Perfekt mit *sein* wie bei Zustandsveränderungsverben üblich.

---

1 Primus (2012, S. 25) zieht den Begriff Handlungskontrolle der Dowty'schen Volition vor.

Tabelle 1. Verortung der Konstruktion zwischen Transitivität und Intransitivität

Transitive Handlung	Unbeabsichtigte Verursachung	Intransitives Geschehen
Agens – Patiens	Verursacher – Patiens	Patiens
NPnom – NPakk	NPdat – NPnom	NPnom
<i>Der Handwerker hat den Krug zerbrochen.</i>	<i><u>Dem</u> Handwerker ist der Krug zerbrochen.</i>	<i>Der Krug ist zerbrochen.</i>
<i>Der Handwerker hat den Krug kaputt gemacht.</i>	<i><u>Dem</u> Handwerker ist der Krug kaputt gegangen.</i>	<i>Der Krug ist kaputt gegangen.</i>
<i>Der Handwerker hat den Krug runtergeworfen.</i>	<i><u>Dem</u> Handwerker ist der Krug runtergefallen.</i>	<i>Der Krug ist runtergefallen.</i>
Perfekt mit <i>haben</i>	Perfekt mit <i>sein</i>	Perfekt mit <i>sein</i>
	Der Verursacher ist negativ betroffen	

## 1.2 Typologische Einordnung

Wie ordnet sich das Deutsche mit dieser Kodierung in die Typologie des Ausdrucks Unbeabsichtigter Verursachung ein? Kittilä (2005, S. 392–394) unterscheidet drei Haupttypen:

1. Transitiver Satz mit zusätzlichem adverbialem Ausdruck wie in *He broke the window accidentally*.
2. Lexikalischer Ausdruck wie in *mislay* oder *verlegen*.
3. Morphosyntaktische Konstruktionen mit vier Untertypen, die sich dadurch unterscheiden, ob sich im Vergleich zum transitiven Satz die Verbmorphologie und / oder die Argument-Markierung verändert; das Verb selbst ist – vs. Typ 2 – neutral.

In einigen Sprachen des Typs 3 erhält das transitive Verb zusätzliche morphologische Markierungen (Subtyp 1), in anderen wird statt des transitiven ein intransitiviertes Verb bei gleichem Kasusrahmen verwendet (Subtyp 3), ein weiterer Subtyp behält das transitive Verb bei, kodiert aber die Argumente, vornehmlich den Verursacher, anders (Subtyp 2), Subtyp 4 schließlich kombiniert die Verwendung des intransitiven Verbs mit einer anderen Kodierung des Verursachers.

Deutsch gehört zu dem vierten, semantisch besonders expliziten Subtyp, indem das intransitive Gegenstück zum transitiven Verb verwendet wird und

gleichzeitig der Verursacher im Dativ erscheint. Wie Tabelle 1 zeigt, ist diese Kodierung jedoch formal unaufwändig – im Vergleich zu den Typen 1 und 2 und zu den zusätzlichen Verbmarkierungen in einigen der Subtypen von 3. Diesen geringen Kodierungsaufwand ermöglicht das im Deutschen noch bestehende Kasussystem, gleichzeitig entstehen dadurch aber Abgrenzungsprobleme zu anderen Konstruktionen mit Dativ.

### 1.3 Eng verwandt: Die *Dative-of-Inaction*-Konstruktion für Nichtverhinderte Ereignisse

Eine formal identische, semantisch aber in Teilen unterschiedliche Konstruktion liegt in Sätzen wie (4) vor.

- (4a) Ihm ist die Suppe übergekocht.
- (4b) Ihm sind die Goldfische eingegangen.

Hens (1997) postuliert für seine Sätze (4) eine *Dative-of-Inaction*-Konstruktion mit folgenden Eigenschaften: Das Patienssubjekt erfährt die vom Verb ausgedrückte Zustandsveränderung, deren Ursache in äußeren Kräften oder der Natur des Subjektreferenten liegt: „The cause for the change lies within the undergoer subject itself; it is understood as a natural process toward which the subject is predisposed“ (Hens, 1997, S. 205). Der Dativrezipient, dem der Subjektreferent gehört oder anvertraut ist oder dem dafür Verantwortlichkeit zukommt, ist wieder hoch agensfähig und wäre in der Lage, den Prozess zu verhindern, handelt aber nicht entsprechend, sondern bleibt inaktiv. Es bietet sich daher an, hier von einer Konstruktion Nichtverhinderter Ereignisse zu sprechen.

Hinsichtlich Dowtys Proto-Rollen-Ansatz liegt in solchen Fällen schwächere Agentivität vor, denn es fehlt die direkte Verursachung des Ereignisses durch den Dativreferenten wie bei Unbeabsichtigter Verursachung, notwendig ist aber Besitz<sup>2</sup> oder zumindest unterstellte Verantwortlichkeit für den Subjektreferenten. Beiden Konstruktionen gemeinsam ist die Wahrnehmung des Ereignisses (Sentience / Perception) sowie die Unabsichtlichkeit bzw. fehlende Handlungskontrolle, die sich in (5) zeigt.

- (5) \*Ihm ist die Suppe absichtlich übergekocht.

---

2 Besitz ist keine der Proto-Agens-Eigenschaften, die Dowty ansetzt, wird jedoch von anderen Autoren, u. a. Jackendoff (1990), Zifonun et al. (1997) und Primus (2012) als eine solche angesehen.

Im Fall Nichtverhinderter Ereignisse sind meist bedeutungsähnliche Sätze mit *lassen* möglich<sup>3</sup>, nicht bei unbeabsichtigt verursachten (vgl. Eisenberg, 2020, S. 319):

- (6a) Er hat die Suppe überkochen lassen.
- (6b) Er hat die Goldfische eingehen lassen.

Wie verhalten sich die von Kittilä beziehungsweise Hens postulierten Konstruktionen zueinander? Oder liegt nur eine – polyseme – Konstruktion vor? Hens (1997, S. 207) subsumiert Sätze wie *Mir ist die Fensterscheibe zerbrochen* als einen der „less typical cases“ unter seinen *Dative-of-Inaction*, was angesichts der hier vorliegenden direkten Verursachung den für ihn zentralen Faktor Inaktivität überzieht. Kittilä (2005) dagegen sieht die Konstruktion der Unbeabsichtigten Verursachung als zentral an und ordnet den Ausdruck von Ereignissen wie *Mir ist mein Hund gestorben* als zusätzliche, in einigen Sprachen mit der zentralen Konstruktion verbundene Funktion ein. In anderen Sprachen verbinden sich mit der zentralen Konstruktion weitere Funktionen wie Ausdruck von Ereignissen, die gegen alle Erwartung stattfinden (Japanisch, Bagwalal), und Ausdruck letztendlichen Gelingens (Thompson River Salish, Kammu, Finnisch).<sup>4</sup> Kittilä (2005, S. 411–414) spricht daher vorsichtiger von „anderen Konstruktionen“.

Wenn Argumentstruktur-Konstruktionen grundlegende menschliche Erfahrungen kodieren (vgl. Goldberg, 1995, S. 39), ist von unterschiedlichen Konstruktionen auszugehen. Dafür sprechen die Unterschiede im Grad der Direktheit der Verursachung: Im einen Fall besteht zeitliche und räumliche Kontiguität zwischen der Verursachung durch den Dativreferenten und der sich am Subjekt vollziehenden Zustandsveränderung, sein Tun ist benennbar, während bei nichtverhinderten Ereignissen der Subjektreferent ohnehin einer Veränderung unterworfen ist (Suppe auf einer heißen Herdplatte kocht irgendwann über, Goldfische im Aquarium gehen ohne Versorgung ein) und der als verantwortlich dargestellte Dativreferent in der Veränderungssituation nicht einmal zugegen sein muss. Was er wann tut oder nicht tut, wird in der Konstruktion nicht ausgedrückt.

Dass Kittilä von den beiden Konstruktionen die Unbeabsichtigte Verursachung als die zentralere ansieht, begründet er durch seinen typologischen Überblick. Auch in didaktischer Hinsicht bietet diese Konstruktion einen geeigneten Ausgangspunkt: Sie ist nah an der den Lernern bekannten transitiven

---

3 Der ausgesprochenen Polysemie von *lassen* ist allerdings geschuldet, dass (6) auch als kausativ und das Ereignis damit als beabsichtigt gelesen werden kann: *Er hat die Suppe absichtlich überkochen lassen*.

4 Zu dieser weiteren verwandten Konstruktion siehe Abschnitt 1.4.



Konstruktion, von der sie sich lediglich durch ein Merkmal unterscheidet, die gut fassbare Unabsichtlichkeit. Dagegen lässt sich bei Nichtverhinderten Ereignissen nur schwer ein korrespondierender transitiver Satz finden: Das (Nicht)Tun des Dativreferenten ist gestaltloser und daher schwerer fassbar. Schließlich ist es aus kognitiver und kommunikativer Sicht wichtig, die Art von Verantwortlichkeit bei Unbeabsichtigter Verursachung verstehen und ausdrücken zu können.

Wegen der Gemeinsamkeiten beider Konstruktionen (gleiche syntaktische Form, kein Vorliegen von Intentionalität, Negativität) mag es in einigen Fällen vom Kontext abhängen, ob ein Satz im Sinne unbeabsichtigter Verursachung oder eines nicht verhinderten Ereignisses gelesen wird. Nach dem Grad der Agentivität lassen sich beide Konstruktionen auf einer gemeinsamen Skala anordnen, die nach unten noch durch den Dativus incommodi wie in (3b) zu ergänzen ist. Auf der obersten Stufe ist der Dativreferent zentraler Mitspieler, der unbeabsichtigt das tut, was das kausative Gegenstück zum intransitiven Verb der Konstruktion beschreibt (*zerbrechen, kaputtmachen, runterwerfen*). Beim *Dative-of-Inaction* ist der Subjektreferent auch ohne einen weiteren Partizipanten Veränderungen unterworfen, der Dativreferent wirkt hier nur insofern mit, als er diese Ereignisse nicht verhindert. Für diesen Fall trifft Wegeners (1985, S. 317) Beschreibung zu, dass offenbleibt, was er tatsächlich getan hat oder ob er nichts getan hat. Am unteren Ende der Skala steht der Ausdruck von Ereignissen, an denen der Dativreferent gar nicht mehr mitwirkt, sondern deren Zeuge und Betroffener er nur noch ist wie in (3b), was dem traditionellen Dativus commodi/incommodi entspricht.<sup>5</sup>

Trotz der unterschiedlichen Agentivität des Dativreferenten ist den drei Ereignistypen gemeinsam, dass sie sich durch das generelle Verb *passieren*

---

5 Die IDS-Grammatik setzt für alle drei Fälle dieselbe Argumentstruktur mit der semantischen Rolle Benefaktiv an. Von dieser „kann bei ergativen Verben in bestimmten Zusammenhängen auf eine bestimmte Form initiiert oder nicht verhindernder Beteiligung im Vorfeld des Ereignisses nur geschlossen werden.“ (Zifonun et al., 1997, S. 1343) So liege nur ein zusätzlicher interpretativer Effekt auf der Basis einer typischen Dativrolle vor. Diese Analyse scheint durch das holistische Rollenkonzept bedingt, wonach Dativkomplementen nur Rezipient, Possessor, Experiencer oder Benefaktiv zugeordnet werden kann, während hier mit Dowty von Proto-Agens-Dimensionen ausgegangen wird, was differenziertere Analysen ermöglicht. Auch angesichts des typologischen Überblicks Kittiläs sollte von eigenständigen Konstruktionen ausgegangen werden. Unplausibel schließlich ist in Bezug auf Sätze wie (1) die Analyse, „daß der durch das periphere Dativkomplement Genannte kein Ereignisbeteiligter im engeren Sinne ist. An dem Vorgang des Zerbrechens einer Vase selbst ist die Person – anders als bei der Handlung ‚Person zerbricht Vase‘ – nicht beteiligt ...“ (Zifonun et al., 1997, S. 1343).

als Quasi-Hyperonym<sup>6</sup> beschreiben lassen: (1') *Das Zerschneiden der Vase ist dem Besucher passiert*; (4a') *Das Überkochen der Suppe ist ihm passiert*; (3b') *Das Vollaufen des Kellers mit Wasser ist den Leuten passiert*.

#### 1.4 Die Komponenten Dativ und „ergative Verben“

Auf der Formseite der Konstruktion sind der Dativ und die Art des Verbs zu besprechen. Zunächst zu den Kasus: Der Nominativ wird von dem intransitiven Verb in die Konstruktion der Unbeabsichtigten Verursachung eingebracht und kodiert das Patiens, an dem sich eine Zustandsveränderung vollzieht. Der Dativ wird für den weiteren Beteiligten gewählt, weil genügend Proto-Agens-Eigenschaften vorliegen: Verursachung, gegebenenfalls Bewegung hinsichtlich des Nominativarguments, Wahrnehmung des Ereignisses; lediglich Volition bzw. Kontrolle fehlt, was die Verwendung des Dativs deutlich signalisiert. In der *Dative-of-Inaction*-Konstruktion liegen weniger Proto-Agens-Eigenschaften vor, da die Verursachung indirekt und nur mitwirkend ist und Bewegung hinsichtlich des Nominativarguments offenbleibt, dafür liegen aber Besitz oder Anvertrautheit vor. Noch einmal weniger Proto-Agens-Eigenschaften weist der Dativ in Konstruktionen mit dem (In)Commodi auf. In allen drei Fällen signalisiert der Dativ, dass trotz eines prinzipiell handlungsfähigen Partizipanten „die Handlungskontrolle über das vom Verb bezeichnete Ereignis ausgeschlossen ist“ (Primus, 2012, S. 59).

Die in der Konstruktion erscheinenden Verben werden oft ergative Verben genannt. Sie bilden eine Untergruppe der intransitiven Verben mit der schon genannten Eigenschaft, dass sie das Perfekt mit *sein* bilden; ihr Partizip II kann attributiv verwendet werden: *das eingeschlafene Kind* (← *das Kind ist eingeschlafen*), aber \**das geschlafene Kind* (← *das Kind hat geschlafen*).

Der Begriff ergativ stammt aus der Typologie, wo er den Kasus bezeichnet, der das Agens typischer transitiver Sätze kodiert. Das Patiens erscheint im morphologisch meist geringer markierten Kasus Absolutiv (auch Nominativ genannt), der auch den einzigen Partizipanten einstelliger intransitiver Verben kodiert, unabhängig davon, ob ein Agens oder Patiens vorliegt. Der Absolutiv ist es auch, der bei Kongruenz am Verb markiert wird; in der Grundabfolge folgt er dem Ergativ. Damit verhalten sich Ergativsprachen komplementär zu Nominativsprachen, wo das Agens transitiver Sätze und der einzige Partizipant einstelliger intransitiver Sätze im selben Kasus erscheinen, bei Kongruenz mit dem Verb korrespondieren und in der Regel morphologisch weniger

---

6 Lyons (1977, S. 298–9) spricht bei Verben wie *act, move, become, make ...* von „very general verbs“ und bezeichnet deren Beziehung zu konkreteren Verben als „quasi-hyponymy“.

aufwändig kodiert werden als der Akkusativ, der nur das Patiens in transitiven Sätzen ausdrückt.

Ein Satz wie *Dem Handwerker ist der Krug zerbrochen* erinnert insofern an die Verhältnisse in Ergativsprachen, als das Patiens im neutralsten Kasus Nominativ erscheint und mit dem Verb kongruiert, während der agentivere Verursacher morphologisch aufwändiger im Dativ erscheint und in der Grundabfolge vorangeht. Im Unterschied zu den Verhältnissen in Ergativsprachen kommen dem Dativpartizipanten mit dem Fehlen der Absicht aber nicht alle Proto-Agens-Eigenschaften zu. Interessanterweise verwenden Ergativsprachen, sofern sie über die Konstruktion Unbeabsichtigter Verursachung verfügen, in solchen Fällen gerade nicht ihren Ergativ, sondern wie das Lesgische einen Adelativ (Haspelmath, 1993) oder sogar – parallel zum Deutschen – den Dativ. Primus (2012, S. 37) merkt daher an: „Dass manche Forscher Konstruktionen wie ... [*Heute ist dem Opa eine teure Vase zerbrochen*] als *ergativ* bezeichnen, ist daher aus typologischer Perspektive irreführend“, und präferiert den Terminus *unakkusativ*. Wegen der größeren Verbreitung in Grammatiken des Deutschen soll hier aber weiter von *ergativ* die Rede sein.

Abschließend sei noch darauf hingewiesen, dass Verben wie *gelingen*, *misslingen*, *glücken*, *missraten* usw., die ebenfalls den ergativen Verben zugeordnet werden (vgl. Eisenberg, 2020), bei weitgehend gleichem syntaktischem Verhalten – sie sind aber zweistellig! – als Lexeme eine andere Semantik mitbringen: Sie signalisieren die volle Absicht, eine Handlung auszuführen, deren Auslösung auch unter Kontrolle ist – es fehlt jedoch die durchgängige Kontrolle über deren Ausführung und damit eine Proto-Agens-Eigenschaft, was im Deutschen wie beim Fehlen der Kontrolle über die Auslösung des Geschehens zu einer Dativkodierung führt.

(7) Mir ist der Kuchen / die Reparatur gelungen / missglückt.

## 1.5 Eine Konstruktion im Sinne der Konstruktionsgrammatik

Kittilä (2005) spricht durchgängig von Konstruktion, ohne sich explizit auf die Konstruktionsgrammatik nach Lakoff (1987), Goldberg (1995) u. a. zu beziehen, ein Bezug, den Hens (1997) dagegen explizit herstellt. Dieses Grammatikmodell nimmt Konstruktionen unterschiedlichen Umfangs und Abstraktheitsgrades an. Mit Unbeabsichtigter Verursachung und Nichtverhinderten Ereignissen liegen verwandte Argumentstruktur-Konstruktionen vor, in denen die formale Struktur ungeachtet ihrer lexikalischen Füllung – was Bedingungen für letztere nicht ausschließt – einen eigenen Beitrag zur Bedeutung leistet.

Sowohl die spezifische Interpretation des Dativs als Verursacher bzw. Verantwortlicher wie auch die negative Bewertung des Ereignisses erfüllen bereits die erste, semantische Bedingung der folgenden Konstruktions-Definition

von Goldberg (2006), deren zweiter Teil die bei Goldberg (1995) noch rein semantische zu einer kognitiven Definition erweitert:

Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency. (Goldberg, 2006, S. 5)

Außerhalb der beiden hier besprochenen Konstruktionen gilt: Tritt ein Dativ im Rahmen einer Transitivkonstruktion mit Agens und Patiens auf und liegt kein Rezipient vor, so ist seine Interpretation als Nutznießer oder Geschädigter (Dativus commodi bzw. incommodi) von Weltwissen und Kontext abhängig (*Peter fuhr seinen Eltern das Auto in die Garage vs. in das Blumenbeet*) und in jedem Kommunikationsakt immer wieder erneut herzustellen, während bei Unbeabsichtigter Verursachung bzw. Nichtverhinderten Ereignissen die Negativität durch die Konstruktion gegeben ist. Das schließt Motiviertheit von Negativität (vgl. 1.1) und Dativwahl (hohe Zahl von Proto-Agens-Eigenschaften, vgl. 1.4) nicht aus, weshalb Hens (1997, S. 215) zu widersprechen ist:

... we need not worry about the unusual dative case assignment for the agent. The indirect object function—and with it the dative case marking—are prelinked to the role in the construction. They are inherent and thus need not be accounted for by more general linking rules, such as those that assign nominative to subjects and accusative to direct objects.

Festzuhalten ist mit Hens aber, dass der Dativ von der Konstruktion, nicht vom Verb eingebracht wird, wovon auch Kittilä für Unbeabsichtigte Verursachung ausgeht. Der Nominativ wird von dem Verb eingeführt, welches den Kern der in die Gesamtkonstruktion eingebetteten Intransitiv-Konstruktion bildet.

Die Einführung des Dativarguments durch die Konstruktion vermeidet Schwierigkeiten, auf die eine verbvalenzbasierte Analyse stößt: Ergative Verben wie *herunterfallen*, *ausgehen* usw. können mit und ohne Dativ auftreten. Da keine freie Angabe vorliegt, wie Wegener (1985) ausführlich begründet hat, müsste es sich um fakultative Ergänzungen handeln. Doch während weggelassene Objekte wie in *Ich schenke gerne Bücher* oder *Heute essen wir schon früher* in der vom Verb evozierten Szene angelegt sind, etabliert der Dativ bei ergativen Verben wie den obigen eine gänzlich andere Szene. So müssten für alle diese Verben zwei Valenzmuster angesetzt werden, neben dem einstelligen ein zweistelliges mit zusätzlicher Dativergänzung.

Im Rahmen der Konstruktionsgrammatik wird davon Abstand genommen, die Valenz der Verben derart auszuweiten. Einer valenzgrammatisch orientierten Beschreibung entgeht auch die systematische Stellung der Konstruktion Unbeabsichtigter Verursachung zwischen der Transitiv- und der Intransitiv-Konstruktion, die Tabelle 1 zeigt. In deren mittlerer Spalte sind die Form- und Inhaltseigenschaften – einschließlich der Negativität – aufgeführt, die gemeinsam die Konstruktion ausmachen.

## 2 Zum Lernproblem

### 2.1 Die Konstruktion in der Didaktik des Deutschen als Fremdsprache

Während in der linguistischen Forschung und in wissenschaftlichen Grammatiken des Deutschen die Funktionen des Dativs und Konstruktionen mit ergativen Verben Gegenstand ausführlicher Diskussion sind, finden sie in pädagogischen, an Lehrende gerichteten Grammatiken trotz ihres umfassenden Beschreibungsanspruchs keine oder nur knappe Erwähnung.

Von den älteren Standardgrammatiken für Deutsch als Fremdsprache geben Schulz & Griesbach (<sup>1</sup>1995) keine Auskünfte. Helbig & Buscha (2013, S. 263) widmen dem Thema fünf Zeilen, in denen der „Dativus incommodi (Dativ des Gelingens / Misslingens, der Verantwortlichkeit)“ als eine der Satzgliedfunktionen des Dativs mit drei Beispielen aufgeführt wird: *Dem Gärtner sind die Blumen verwelkt, Der Schlüssel ist dem Kind ins Wasser gefallen, Das Kind zerbrach den Eltern die Vase*. In einer Anmerkung definieren die Autoren den Incommodi als eine Person bezeichnend, welcher der Subjektreferent anvertraut war und „der das Geschehen als negativ, nicht-intentional und unerwünscht erscheint. Es kann sich dabei um Tätigkeiten oder Vorgänge handeln, der Dativ-Referent – dem das Geschehen ‚passiert‘ – ist kein intentionales Agens“ (Helbig & Buscha, 2013, S. 265). Hier werden Aspekte der Konstruktion genannt, aber die Subsumierung unter Dativus incommodi ist unglücklich, nicht nur weil dieser auch als Dativ des Gelingens (!) charakterisiert wird, sondern auch, weil er üblicherweise zusammen mit dem Dativus commodi und als dessen Antonym definiert wird, als Dativ des Geschädigten im Gegensatz zu dem des Nutznießers. Auch bleibt eine genaue Abgrenzung des Agentivitätsgrads dieses Dativs im Verhältnis zu anderen Dativen offen.

Von neueren pädagogischen Grammatiken für die Vermittlung von Deutsch als Fremd- und Zweitsprache gibt auch L. Hoffmann (<sup>3</sup>2016) keine Auskünfte. Fandrych & Thurmair (<sup>2</sup>2021, S. 220) führen *Die teure Vase ist mir leider heruntergefallen* unter den sog. freien Dativen auf und charakterisieren die Funktion des Dativs wie schon Helbig & Buscha als Incommodi, ohne dass

der Begriff erläutert oder von einem nicht-intentionalen Agens gesprochen würde; nahegelegt wird damit lediglich negative Betroffenheit.

Lerner- und Übungsgrammatiken thematisieren weder die Konstruktion der Unbeabsichtigten Verursachung noch die verwandte Konstruktion Nicht-verhinderter Ereignisse. Das gilt für ältere wie aktuelle, für Grundstufen- wie Fortgeschrittenengrammatiken.<sup>7</sup> In der Regel enthalten sie in ihrem Anhang lediglich Listen von zweistelligen Verben mit Dativ, die in ihrer alphabetischen Anordnung aber keinen Unterschied machen zwischen Verben mit einem Agenssubjekt und einem Dativobjekt wie *antworten, danken, helfen, folgen, vertrauen, zustimmen* und den sich gänzlich anders verhaltenden Verben mit dem agentiveren Argument im Dativ und einem Patienssubjekt wie Empfindungs- und Wahrnehmungsverben (*gefallen bzw. auffallen*) oder Verben des Gelingens / Misslingens sowie weiteren Gruppen von Dativverben;<sup>8</sup> *geschehen* und *passieren* erscheinen in den Listen nicht, weil aus valenzgrammatischer Sicht, welcher die Lehrmaterialien überwiegend folgen, kein obligatorischer Dativ vorliegt.

Der Erwerb von Konstruktionen mit einem Dativ, der das agentivste Argument des Satzes kodiert, wird offenbar dem impliziten Lernen über die Verblexeme überlassen. Dieser Weg ist bei der Konstruktion der Unbeabsichtigten Verursachung aber nicht gangbar, weil die hier auftretenden intransitiven Verben – etwa im Gegensatz zu *gelingen* usw. – als Lexeme keinen Hinweis auf die Semantik des Dativs geben. So zeigen Lernerdaten, dass sich die Konstruktion auch bei Fortgeschrittenen dem impliziten Erwerb weitgehend entzieht.

## 2.2 Einige Lernerdaten und ihre Interpretation

Die folgenden Daten wurden in einem Fortbildungskurs für Lehrer des Deutschen als Fremdsprache erhoben. Die zwölf Teilnehmer waren durchschnittlich 31 Jahre alt und unterrichteten Deutsch meist schon einige bis mehrere Jahre lang; ihre Ausgangssprachen waren Albanisch, Arabisch, Englisch, Französisch, Griechisch, Hindi, Italienisch, Koreanisch, Persisch und Tschechisch.

---

7 Vgl. nach aufsteigenden Niveaustufen Kars & Häussermann: *Grundgrammatik Deutsch* (1988), Reimann: *Grundstufen-Grammatik* (1996), Jin & Voß: *Grammatik aktiv A1–B1* (2013), Fandrych & Tallowitz: *Klipp und Klar – Neubearbeitung* (2021), Hohmann & Rohrmann *Grammatik mal vier* (2021), Latour: *Mittelstufen-Grammatik* (1988), Dreyer & Schmitt: *Lehr- und Übungsbuch der deutschen Grammatik – aktuell* (2009), Jin & Voß: *Grammatik aktiv B2–C1* (2017), Hall & Scheiner: *Deutsch – Übungsgrammatik für die Oberstufe – aktuell* (2020).

8 Beispiele aus der Liste „Verben mit Dativ“ in Hohmann & Rohrmann (2021): *Grammatik mal vier*.

Das Material bestand aus einem Kurzfilm mit dem Titel „Date“<sup>9</sup>, der in einigen Ausschnitten gezeigt wurde. Im ersten verabredet sich ein junger Mann am Telefon mit einer Frau zu einem Date. Nach dem Auflegen trägt er den Termin ein, wobei er kurz zögert, denn es ist ein Freitag, der 13. Hier wurde der Film unterbrochen, um zu klären, welche Assoziationen sich im deutschsprachigen Raum mit dem Datum verbinden, womit auch der Rahmen für die Erwartung negativer Ereignisse gesetzt wurde. Ein zweiter Ausschnitt lässt den Zuschauer Zeuge der bemühten Vorbereitungen des Mannes am Vorabend des Dates werden, während der dritte Ausschnitt all die kleineren und größeren Unglücke zwischen seinem Aufwachen und dem Verlassen des Hauses zeigt. Diese gut fünfminütige Sequenz war Basis der Datenerhebung. Nach dem ersten Sehen erhielten die Teilnehmer ein Arbeitsblatt mit Standbildern von neun Missgeschicken der Sequenz. Die Formulierung der Aufgabe, die Szenen mit den Standbildern zu beschreiben, war so gehalten, dass die Verwendung der Konstruktion der Unbeabsichtigten Verursachung nicht suggeriert wurde (*Was war los? Beschreiben Sie diese Szenen aus dem Film, wo der Mann sich auf das Date vorbereitet hat*).<sup>10</sup> Die Standbilder griffen folgende Szenen der Sequenz heraus:

- I. Beim Ausschalten des Weckers fällt dem jungen Mann das Nadelkissen herunter, das er am Vorabend zum Annähen eines Knopfes gebraucht hatte. Beim Aufstehen tritt er in die Nadeln und verletzt sich am Fuß.
- II. In der Dusche springt ihm die Seife aus der Hand, über die er später noch fallen wird.
- III. Er verletzt sich die Nase mit dem Nasenhaartrimmer.
- IV. Ihm bleibt die Watte des Wattestäbchens, mit dem er das Blut stillt, in der Nase stecken.
- V. Beim Frühstück verbrennt ihm der Toast und
- VI. er gießt sich schlecht gewordene Milch in den Kaffee.
- VII. Als er sich den Mund spülen will, muss er feststellen, dass das Leitungswasser abgestellt wurde.
- VIII. Auf der Suche nach Mineralwasser als Ersatz findet er nur leere Flaschen.
- IX. In seiner Verzweiflung will er sich den Mund mit dem Wasser einer Blumenvase ausspülen, die er aber umstößt.

---

9 Quelle <https://vimeo.com/3349493>, zuletzt abgerufen am 19.09.2022.

10 In einem anderen Lehrerfortbildungskurs mit 19 Teilnehmern lautete die Aufgabe so: *Was ist dem Mann passiert, als er sich auf das Date vorbereitet hat? Beschreiben Sie diese Szenen aus dem Film*. Auch diese Formulierung führte trotz möglichem syntaktischem Primingeffekt nicht zu einer breiteren Verwendung der Konstruktion.

Vor der schriftlichen Beschreibung der Szenen sahen die Teilnehmer die Sequenz ein zweites Mal.

Allgemein zeigt die Auswertung, dass die Konstruktion nur von wenigen Teilnehmern für zwei der neun Szenen spontan verwendet wird; daneben werden andere passende Versprachlichungsmittel eingesetzt, aber auch unangemessene agentive Beschreibungen mit einem transitiven Satz. Im Einzelnen: Die Konstruktion erscheint in (8)–(9),

- (8) Die Seife glitt ihm aus den Händen auf den feuchten Fliesen.  
Ihm ist die Seife aus der Hand gerutscht.  
Beim Duschen rutschte ihm die Seife aus seiner Hand über die Duschwanne auf den Fliesenboden hinaus.
- (9) Ihm ist die Vase aus der Hand gerutscht.

andere Teilnehmer drücken die Unabsichtlichkeit adverbial aus,

- (10a) Der Mann hat den Wecker ausgeschaltet und hat dabei den Nadelkissen aus Versehen auf den Boden geworfen. Damit begann sein Freitag, der 13.
- (10b) Erstmal hat er versucht, den Wecker auszuschalten. Unabsichtlich hat er das Nadelkissen auf den Boden gefällt.

zwei Teilnehmer verwenden *lassen*,

- (11a) Er hat seine Toast anbrennen lassen.
- (11b) Er hat die Vase fallen lassen ...

wieder andere signalisieren die Sachverhalte durch den Textaufbau, indem sie eine agentive Handlung des Mannes eng mit dem intransitiv formulierten unbeabsichtigten Ereignis verknüpfen; (12d) kombiniert diese Strategie mit einem Adverbial:

- (12a) Während er sich geduscht hat, ist die Seife aus seinen Händen gerutscht.
- (12b) Um die Blutung zu stoppen, hat er ein Wattestäbchen verwendet. Jedoch ist die Watte in seiner Nase steckengeblieben.
- (12c) Er hat das Brot im Toaster vergessen, daher ist es verbrannt.
- (12d) Er ist früh aufgewachen. Als erstes hat er den R.wecker ausgemacht. Dabei ist das Nadelkissen unglücklicherweise runtergefallen.



Neben der Konstruktion der Unbeabsichtigten Verursachung (8–9) führen also die Strategien Verwendung von Adverbialen (10), *lassen* (11) und Textaufbau (12) zu inhaltlich angemessenen Beschreibungen der Szenen. Wohl nicht zufällig wird die Konstruktion gerade für die Szenen II und IX verwendet. Beide sind insofern prototypisch, als der Protagonist eine zentrifugale Bewegung mit unmittelbarem Kontakt zur Seife bzw. Vase ausführt und das Ereignis direkt auslöst.<sup>11</sup> Andere Szenen wie I sind diesbezüglich weniger prototypisch (beim Ausschalten des Weckers verschiebt der Mann den Wecker, was indirekt zum Herunterfallen des Nadelkissens führt) oder zeigen wie V–VII nicht verhinderte Ereignisse: dass ihm der Toast verbrannt, die Milch schlecht geworden und das Mineralwasser ausgegangen ist.

Als misslungen dagegen müssen aktive Beschreibungen mit transitiven Sätzen wie in (13) gelten,

(13) Er verbrannte die Toaste. / Er warf die Vase um.

wenn nicht umgebender Text wie etwa *Als er das Wasser aus der Vase nehmen wollte, warf er die Vase um* die Intentionalität abschwächt. Vergleichbar unangemessene Verwendungen intransitiver Sätze waren nicht zu beobachten, da sie entweder wie in (12) eng mit einer Handlung verbunden werden und das Ereignis so als deren indirekte Ursache verstanden werden kann oder weil der Protagonist in der zu beschreibenden Szene nicht zwingend als verursachend oder verantwortlich gesehen werden muss, wie in VI, wo *Die Milch ist abgelaufen* eine mögliche Beschreibung abgibt.

Auffällig ist auch die nicht seltene unangemessene Verwendung des Passivs wie in (14) und in (15b).

(14a) Er dachte auch das Wasser der Blumenvase verwenden. Aber die Vase ist umgeworfen worden.

(14b) Er hat die Toasts vergessen und sie wurden verbrannt.

(14c) Auch mit dem Frühstück hat es nicht funktioniert. Das Brot wurde beim Toasten gebrannt.

Es muss offenbleiben, ob hier mit dem Passiv und seiner Defokussierung des Agens und Fokussierung des Vorgangs die schwächere Agentivität des Protagonisten ausgedrückt werden soll und ob dies durch Passiv oder passiv-ähnliche Strukturen der L1 begünstigt wird.

---

11 Auffällig ist, dass alle Dative in der Seifenszene in Kombination mit *aus der Hand*, *aus seiner Hand* oder *aus seinen Händen* auftreten. Hier könnte auch ein Chunk mit Pertinenzdativ vorliegen.

Eine Teilnehmerin mit Koreanisch als L1, von der die korrekt verwendete Konstruktion in (9) stammt, versucht sich fast durchgehend in ihrer Anwendung, hat dabei aber Probleme mit der Transitivität des Verbs wie in (15a–b) oder formuliert einen korrekten Satz wie (15c), der inhaltlich nicht zutrifft, da das Wasser dem jungen Mann von Handwerkern abgestellt wird, oder drückt sich wie in (15d) hinsichtlich der Kausalität unglücklich aus.

- (15a) Ihm hat die Seife aus der Hand gerutscht.
- (15b) Ihm wurde das Wattebällchen in der Nase steckengeblieben.
- (15c) Der Hahn ist ihm kaputtgegangen.
- (15d) Ihm hat der Nasenhaarschneider wehgetan.

Im Nachgang zur Erhebung und späteren Diskussion der obigen Daten wurde nach der Bekanntheit der Konstruktion in Form und Bedeutung gefragt. Die rezeptive Kenntnis war erwartungsgemäß breiter als die aktive Verwendung, von der teilweise wegen Unsicherheiten im Gebrauch zugunsten anderer Ausdrucksweisen Abstand genommen worden war, mit denen die Sachverhalte ja angemessen auszudrücken waren. Einige Teilnehmer zeigten sich selbst erstaunt über ihre Vermeidung der Konstruktion, denn im Albanischen, Italienischen, Griechischen und Tschechischen existiert eine ähnliche, die hätte transferiert werden können. Für Sprecher des Englischen, Persischen oder Hindi besteht keine solche Transfermöglichkeit; diese Sprachen müssen die transitive Konstruktion verwenden, deren höhere Agentivität durch lexikalische Mittel optional abgeschwächt werden kann.

Ein weiterer Faktor, der die durchgehende Verwendung der Konstruktion verhindert haben könnte, ist das Verhältnis von Subjekt und Topik (hier: der junge Mann) in der jeweiligen L1, worauf in 3.1 noch eingegangen wird.

Manche Teilnehmer merkten an, dass ihnen besondere Verwendungen des Dativs jenseits des indirekten Objekts in ditransitiven Sätzen als Lernprobleme bekannt sind wie *mir geht es gut*; *mir ist kalt*; *mir ist bewusst, dass*. Es besteht demnach ein explizites Wissen um Dative, die den Experiencer kodieren, aber weniger ein Wissen um die noch agentiveren Dative in der Konstruktion der Unbeabsichtigten Verursachung, die sich einer Interpretation als bloß negativ Betroffene im Sinne des *Incommodi* anbieten, ein Missverständnis, das nur selten bewusst werden dürfte. Die merkmalsärmere Konzeptualisierung durch Fremdsprachler in der Rezeption bliebe dann weitgehend unentdeckt.

Was entgeht L2-Sprechern, wenn ihnen die Konstruktion der Unbeabsichtigten Verursachung fehlt? Es entgeht ihnen weit mehr als nur ein ökonomisches Ausdrucksmittel für eine bestimmte Ereigniskonstellation, vielmehr ist davon auszugehen, dass sie solche Ereignisse anders konzeptualisieren. In zahlreichen Untersuchungen hat von Stutterheim gezeigt: „Grammatikalisierte

Strukturen lassen sich unter kognitionswissenschaftlicher Sicht als *Filter der Aufmerksamkeitsteuerung* verstehen“ (von Stutterheim & Carroll, 2007, S. 36). Dass auch der grammatikalisierte Ausdruck des Agentivitätsgrads einen solchen Einfluss auf die Kognition von Ereignissen hat, belegen Fausey & Boroditsky (2011) in ihrer Untersuchung zur unterschiedlichen Kodierung unbeabsichtigter Ereignisse durch Englisch- vs. Spanischsprachige. Erstere kodieren durch die Verwendung transitiver Sätze agentiver, letztere wegen einer präferierten Reflexivkonstruktion weniger agentiv. Beide Sprechergruppen weisen signifikante Unterschiede in der Erinnerung an die Verursacher auf. Ferner konnten Fausey & Boroditsky (2010) zeigen, dass agentive und weniger agentive Kodierungen in derselben Sprache (transitive vs. intransitive Formulierungen im Englischen) beeinflussen, welches Maß an Schuld dem Verursacher bei Unabsichtlichkeit zugeschrieben und in welcher Höhe gegebenenfalls Schadensersatz gefordert wird. So lässt sich auch für die deutsche Konstruktion annehmen, dass für L1-Sprecher unbeabsichtigte Verursachung „durch die Grammatik fokussiert“ (von Stutterheim & Carroll, 2007) deutlich von prototypischen (intentionalen) Handlungen abgehoben ist, eine Profilierung, die vielen L2-Sprechern – so lassen die Daten vermuten – nicht automatisiert zur Verfügung steht.

Zusammenfassend: Die Konstruktion der Unbeabsichtigten Verursachung wird Lernenden von Deutsch als Fremdsprache nicht systematisch vermittelt, was auch für die Konstruktion Nichtverhinderter Ereignisse gilt. Weil für die zentrale Interpretation der Rolle des Dativreferenten keine lexikalischen Hinweise des Verbs vorliegen und Lernende zu einer unterspezifizierten Interpretation als bloß negativ Betroffenenem greifen können, entziehen sich beide Konstruktionen und die damit einhergehenden Konzeptualisierungen leicht dem impliziten Erwerb, wie die obigen Daten zeigen. Auf implizites Lernen kann wohl nur gesetzt werden, wenn Deutsch früh in zweitsprachlicher Umgebung erworben wird, wo Entschuldigungen wie *Tut mir leid, das ist mir runtergefallen / Oh, das ist mir kaputt gegangen* und ähnliche Verwendungen der Konstruktion in Alltagssituationen üblich sind. Für Deutsch als Fremdsprache dagegen gilt: Statt allein auf gelegentliche Begegnungen im Input zu setzen, bedarf es didaktischer Intervention in Form von speziell aufbereitetem und kontextualisiertem Input, in aller Regel begleitet von Bewusstmachung mit variablem Explizitheitsgrad (3.2).

### 3 Zur Vermittlung

#### 3.1 Inputgestaltung an einem Beispiel

Konstruktionen bedürfen als konventionalisierte Form-Bedeutungs-Paare zu ihrem Erwerb wiederholten Auftretens im Input. Im Folgenden wird ein Vorschlag unterbreitet, wie Input gezielt angereichert werden kann, um die kognitive Verfestigung der Konstruktionen der Unbeabsichtigten Verursachung und der Nichtverhinderten Ereignisse anzubahnen, und welche Möglichkeiten bewusster Aufmerksamkeitssteuerung bestehen.

Als Material wird wieder der Kurzfilm „Date“ verwendet, der in 2.2 als Mittel der Datenerhebung vorgestellt wurde. Die ersten Phasen des Unterrichts sind identisch: 1. Die Lernenden bestimmen anhand der ersten Sequenz von gut 30 Sekunden die Situation: Ein junger Mann verabredet sich telefonisch mit einer jungen Frau für einen Freitag, den 13., zu ihrem ersten Treffen. 2. Mit einem Assoziogramm wird geklärt, dass sich im deutschsprachigen Raum mit dem Datum die Erwartung unglücklicher Ereignisse verbindet; damit wird die Negativität der Konstruktion situiert. 3. Die Lernenden sehen die für die Vermittlung der Konstruktion zentrale fünfminütige Sequenz mit all den Unglücken, die dem Mann zwischen Aufstehen und Verlassen des Hauses geschehen, zum ersten Mal. Der Ton dieser Sequenz wie des gesamten Videos besteht überwiegend aus Musik und Geräuschen, das wenige Gesprochene hat mit einer Ausnahme Hintergrundcharakter. Die Inhalte erschließen sich also über nicht-sprachliche Informationen. Das macht es möglich, für die nächste Phase gezielt angereicherten sprachlichen Input einzuführen.


In dieser 4. Phase erhalten die Lernenden vor dem zweiten Sehen Vorgaben für die gelenkte kommunikative Aufgabe, die Geschehnisse zu rekonstruieren und schriftlich nachzuerzählen. Ein Arbeitsblatt (Abbildung 1) gibt für sieben zentrale Szenen Subjekt und ergatives Verb vor (z. B. *das Nadelkissen – auf den Boden fallen*). Zweitens fordert es auf, die Struktur *dem jungen Mann / ihm ist ... Verb im Partizip II* zu verwenden und gibt drittens den Textanfang vor: *Das erste, was ihm passiert ist, war im Schlafzimmer. Als er den Wecker ausstellen wollte, ist ihm das Nadelkissen auf den Boden gefallen. Er ist mit seinem Fuß hineingetreten und hat sich verletzt. – Im Bad ist ihm ...*

Die Zielkonstruktionen müssen so mindestens sieben Mal verwendet werden, wobei die Lernenden inhaltlich durch zweimaliges Sehen der Sequenz und sprachlich durch die Vorgaben der Satzteile entlastet sind, so dass sie ihre Aufmerksamkeit auf kommunikative Aufgabe und Konstruktion richten können. Mechanisches Abarbeiten der Vorgaben ist nicht möglich, denn diese sind zeitlich ungeordnet, außerdem sind die Handlungsorte zu berücksichtigen.

Um die Inputflut zu erhöhen, erscheinen auf dem Arbeitsblatt zusätzlich vier Sätze mit dem Verb *passieren*, in der Überschrift (*Freitag, der 13. – Der*

**Freitag, der 13.**  
**Der junge Mann freut sich auf sein Date, doch an diesem Morgen ist ihm einiges passiert.**

die Orte
im Schlafzimmer
im Bad
in der Küche



die Geschehnisse / was passiert ist
das Mineralwasser – ausgehen
das Nadelkissen – auf den Boden fallen
der Toast – verbrennen
die Milch – schlecht werden
die Seife – aus der Hand und aus der Dusche springen
die Vase – umfallen
die Watte – in der Nase stecken bleiben

Erzählen Sie der Reihe nach, was dem jungen Mann in seiner Wohnung passiert ist. Benutzen Sie dafür unter anderem die Satzteile oben in dieser Struktur: *dem jungen Mann / ihm ... Verb im Partizip II.*

Wenn Sie etwas, was ihm geschehen ist, besonders hervorheben wollen, leiten Sie es so ein:  
*Und dann ist ihm noch etwas passiert: ...*

*Das erste, was ihm passiert ist, war im Schlafzimmer. Als er den Wecker ausstellen wollte, **ist ihm** das Nadelkissen auf den Boden **gefallen**. Er ist mit seinem Fuß hineingetreten und hat sich verletzt.*

*Im Bad ist ihm* .....

.....

.....

.....

Abbildung 1. Arbeitsblatt mit kritischem Input

*junge Mann freut sich auf sein Date, doch an diesem Morgen ist ihm einiges passiert*), in der Aufgabenstellung, im Textanfang und als Angebot für die Strukturierung des Textes (*Wenn Sie etwas, was ihm geschehen ist, besonders hervorheben wollen, leiten Sie es so ein: Und dann ist ihm noch etwas passiert: ...*). Die Vor- und Nachteile dieser Gestaltung sollen im Folgenden kurz dargelegt werden.

Syntaktisch verhält sich *passieren* wie die Verben in den fokussierten Konstruktionen; mit seinem Einbezug weist der Input elf ergative Verben in schiefer Verteilung auf: 4 × *passieren*, 2 × (*um*)*fallen*, je 1 × *schlecht werden*, *ausgehen*, *springen*, *stecken bleiben*, *verbrennen*. Schiefe Verteilung ist lernförderlich; sie wurde in an Kinder gerichtetem Input gefunden (Goldberg, 2006, S. 76) und hat sich auch in Experimenten mit Kunstsprachen als am lernwirksamsten erwiesen (Goldberg, 2006, S. 82). Eine solche Verteilung „sollte sowohl zu

gutem Entrenchment des zentralen Typen führen, da dieser häufig wiederholt wird und relativ prominent ist, als auch zu guter Mustererkennung auf Basis der Variation ...“ (Madlener & Behrens, 2015, §3) – Für *passieren* als Anker sprechen auch die Frequenz im Vergleich zu bedeutungsähnlichen Verben<sup>12</sup> und seine Kookkurrenzen: Die 20 häufigsten Subjekte im Kernkorpus 21 des DWDS (Digitales Wörterbuch der deutschen Sprache) drücken Negatives aus (*Schlimm, Unglück, Unfall, Missgeschick, Malheur, Panne, Schrecklich, Fehler*), Unerwartetes (*Unvorhergesehen, Außergewöhnlich, Wunder, Unerwartet, Besondere, Ungewöhnlich*) oder vergleichen das Ereignis mit anderen Ereignissen (*Ähnlich, Gleiche, Gegenteil*).<sup>13</sup>

Für ein genaues Verständnis der Inhaltsseite der Konstruktionen reicht *passieren* jedoch nicht aus, denn das Verb bildet, wie in 1.3 gezeigt, ein Quasi-Hyperonym zu unbeabsichtigt verursachten Ereignissen, nicht verhinderten Ereignissen und schließlich Ereignissen, wo der Dativ nichts Weiteres ausdrückt als eine negativ betroffene Person im Sinne des Incommodi. Für die genaue Bedeutungsprofilierung sind daher bewusstmachende Aufgaben wie in 3.2 notwendig.

Mit der inputinduzierten gehäuftten Verwendung der Konstruktion erschöpfen sich die Vorgaben für die gestellte kommunikative Aufgabe (*Erzählen Sie der Reihe nach, was dem jungen Mann in seiner Wohnung passiert ist*) noch nicht, denn sie lässt sich nicht mit isolierten Sätzen bewältigen, sondern verlangt einen Text, dessen Kohärenzbedingungen sich mit dem Quaestio-Ansatz (Klein & von Stutterheim, 1987) genauer bestimmen lassen: Der Mann wird als globales Topik des Textes gesetzt und für die Zeitdomäne wird vorgegeben, dass die zu verbalisierenden Ereignisse insgesamt vorzeitig sind und untereinander im zeitlichen Nacheinander angeordnet werden. Aus didaktischen Gründen verlangen die Formulierung der Aufgabe und der Textanfang weiter die Verwendung des Perfekts zum Ausdruck der Vorzeitigkeit. Diese Vorgabe ist für Erzählungen nicht zwingend – im Gegenteil ist im konzeptionell Schriftlichen das Präteritum Standard –, wird hier aber gemacht, um das

---

12 Laut Gesamtkorpus des DWDS steigt der Gebrauch von *passieren* seit 1945, während der von *geschehen* sinkt; *widerfahren* ist gegenwärtig 25-mal seltener. Das Goethe-Zertifikat führt *passieren* in der Wortliste für das Niveau A2, *geschehen* für B1.

13 DWDS-Wortprofil für „passieren“, erstellt durch das Digitale Wörterbuch der deutschen Sprache, <<https://www.dwds.de/wp/passieren>>, abgerufen am 03.09.2022. Es kann einem auch etwas Gutes passieren, worauf *Gut* mit Rang 16 der häufigsten Subjekte hinzuweisen scheint. Ein genauerer Blick auf die Belege zeigt aber, dass formelhafte Wendungen mit Negation und *können* dominieren wie in *Etwas Besseres hätte ihr nicht passieren können, ist sie sich heute sicher* oder *Je länger, je mehr könnte man meinen, den Börsen hätte nichts Besseres passieren können als die Corona-Krise*.

mit den ergativen Verben verbundene *sein*-Perfekt zu festigen, ein Lernproblem, welches in einigen der Lernerdaten aufscheint.

Gemäß den Vorgaben ist in einem zielsprachlichen Text für die Sätze der textuellen Hauptstruktur zu erwarten, dass Topikausdrücke wie *dem jungen Mann* oder *ihm* durchgehend im Vorfeld oder direkt hinter dem Finitum *ist* im Mittelfeld stehen – letzteres, wenn im Vorfeld die referentielle Bewegung vorangetrieben wird, durch Adverbiale wie *dann*, *beim Frühstück* für die zeitliche Verschiebung bzw. *in der Dusche*, *in der Küche* für Orts-/Zeitverschiebung. Die jeweils fokussierte Information jedes neuen Missgeschicks erscheint im Subjekt – das als Fokusausdruck dem Topikausdruck im Dativ folgt – und im Partizip II des Verbs in der rechten Klammer.

Über weite Strecken der Hauptstruktur des Textes müssen somit Topik und Subjekt auseinanderfallen. Das ist eine Herausforderung für Lernende mit einer L1, wo die Verbindung von Subjekt und Topik enger oder sogar bindend ist wie im Englischen oder das Subjekt zumindest nicht-pronominalen Objekten vorangehen muss. Bei der Interpretation der Lernerdaten (2.2) wurde bereits vermutet, dass dieser typologische Unterschied zur Vermeidung der Konstruktion beigetragen haben könnte, selbst wenn sie den Lernenden nicht unbekannt ist.

Die Besprechung der Lernertexte muss also nicht nur die formal und inhaltlich korrekte Verwendung der Konstruktion auf Satzebene in den Blick nehmen, sondern auch die Angemessenheit und Kohärenz des Textes, besonders hinsichtlich seiner Topikelemente. Mit dem gegebenenfalls nötigen korrekiven Feedback wird die reine Inputgestaltung überschritten.

### 3.2 Möglichkeiten der Bewusstmachung

Nach der gelenkten Verwendung beider Konstruktionen sind deren jeweilige Bedeutungsseite zu sichern und voneinander abzugrenzen. Dazu werden die sieben Ereignisse noch einmal in Form isolierter, aber vollständiger Sätze in ihrer zeitlichen Abfolge vorgegeben, und die Lerner sollen sie nach folgenden Kriterien ordnen und die Zuordnung begründen:

- Welche Ereignisse hat der Mann selbst verursacht?
- Welche Ereignisse hätte der Mann vielleicht verhindern können?

Das ist eine inhaltliche Aktivität, bei der das im Film Wahrgenommene verarbeitet und diskutiert wird und die so diskursiv wie möglich durchgeführt werden sollte. Ihr Ziel ist, einem groben, unterspezifizierten bloß situativen Verständnis des Dativs als Ausdruck für einen irgendwie negativ Betroffenen entgegenzuwirken und die hohe Agentivität des Verursachers in den einen Situationen und die unterstellte Verantwortlichkeit in den anderen bewusst

zu machen. Sie kann als eine Art von Interpretationsaufgabe angesehen werden, ein Aufgabentyp, der im Rahmen des Formfokussierungsansatzes<sup>14</sup> eine wichtige Zwischenposition einnimmt zwischen angereicherter Input, der implizites Lernen auslöst, und explizitem Regellernen. Interpretationsaufgaben lenken in kontrastivem Vorgehen die Aufmerksamkeit auf Leistungen sprachlicher Formen.

Anspruchsvoller – und je nach Stand der Lerngruppe eventuell zurückzustellen, um Überforderung zu vermeiden – ist die weitere Interpretationsaufgabe, den Unterschied zwischen folgenden beiden Aussagen zu bestimmen:

- Ihm ist die Seife aus der Dusche gesprungen.
- Ihm wurde das Wasser abgestellt. (Nach dem Versuch des Mannes, sich den Mund mit dem Wasser aus der Blumenvase zu spülen, hört man Handwerker rufen „Wasser ist wieder da!“)

Der Unterschied kann von den Lernenden auf der inhaltlichen Ebene beschrieben werden, erwartbar ist aber, dass der zweite Satz als Passivsatz mit weggelassenem Agens (den Handwerkern) bestimmt wird; die Rolle des Dativs ist hier zurecht bloß die des negativ Betroffenen.

Auf der höchsten Stufe der Expliztheit wird die Konstruktion der Unbeabsichtigten Verursachung mit syntaktischen und semantischen Termini dargestellt wie in Tabelle 1. Die Darstellung ist schrittweise zu entwickeln, ausgehend von den vertrauten Konstruktionen für Intransitives Geschehen (rechte Spalte) und transitive Handlung (linke Spalte). Mit beiden ist danach die Zielkonstruktion zu kontrastieren. Innerhalb jeder Spalte sind zuerst die Sätze zu präsentieren, danach ihre syntaktischen Eigenschaften und schließlich ihre semantischen.

Unbeabsichtigte Verursachung sollte zuerst erläutert werden, da sowohl die Verursacherrolle als auch der Ereignistyp in Abgrenzung zu dem transitiven Typ leichter fasslich ist als das Konzept nichtverhinderter Ereignisse, wo das (Nicht)Handeln der als verantwortlich dargestellten Person gestaltloser ist. Diese didaktische Entscheidung wird durch die typologische Untersuchung von Kittilä (2005) gestützt, der Unbeabsichtigte Verursachung als zentralere der beiden Konstruktionen einstuft. In der formalen Darstellung

---

14 Der Ansatz wurde u. a. von Long (1991) und Doughty & Williams (1998) unter dem Schlagwort Focus-on-form eingeführt und ausgearbeitet. In der kognitiv orientierten Lern- und Erwerbstheorie wurzelnd, wendet er sich gegen die weitgehend von Inhalten abgelöste Formenfokussierung (Focus-on-forms), wie sie in der traditionellen Grammatikvermittlung im Fremdsprachenunterricht praktiziert wird, und fordert, statt „auf Vorrat“ die Aufmerksamkeit nur bei Verstehens- und Produktionsproblemen im Rahmen kommunikativer Aktivitäten auf die sprachliche Form zu lenken.



dieser zweiten Konstruktion ist dann „Verursacher“ durch „Verantwortlicher“ zu ersetzen und „Unbeabsichtigte Verursachung“ durch „Nichtverhindertes Ereignis“.

Diese Stufe der Explizitheit ist kein notwendiger Vermittlungsschritt; ob er erforderlich ist, richtet sich nach Alter, formaler Bildung, Lerngewohnheiten und L1 der Lerngruppe, aber auch individuellem Lernstil, der eher holistisch oder analytisch ausgerichtet sein kann. Unabdingbar ist aber angereicherter Input und genaues Erfassen der Inhaltsseite der Konstruktionen, also mindestens Aufmerksamkeit im Sinne der Formfokussierung.

## 4 Schluss

Unbeabsichtigte Verursachung wird im Deutschen durch den noch vorhandenen Dativ höchst ökonomisch, aber auch leicht übersehbar kodiert. Eingebracht über die Konstruktion wird mit diesem Kasus ein weiterer Mitspieler etabliert, was die von dem einstelligen Verb beschriebene Situation der Zustandsveränderung zu einem verursachten Ereignis macht. Das Verb selbst gibt dabei – im Gegensatz zu den ergativen Verben des Gelingens oder Misslingens oder den psychischen Verben mit ihrer zweistelligen Valenz – keinen Hinweis auf die konstruktionsspezifische, aber rollensemantisch motivierte Bedeutung des Dativs. Da Lernenden Dative mit der Bedeutung des Nutznießers oder Geschädigten bekannt sind, bietet sich ihnen eine unterspezifizierte Interpretation als *Incommodi an*, gestützt durch die Negativität, die auch der Konstruktion zueigen ist. Damit wird aber die hohe Agentivität des Dativreferenten im Falle unbeabsichtigter Verursachung bzw. seine Verantwortlichkeit bei nichtverhinderten Ereignissen in der Rezeption verfehlt. In der Produktion tendieren auch L2-Sprecher, deren L1 eine vergleichbare Konstruktion aufweist, dazu, die Konstruktion im Deutschen zugunsten anderer Ausdrucksmöglichkeiten zu vermeiden. Unsicherheit ob der Transferierbarkeit einer als markiert bewerteten L1-Struktur könnte eine Rolle spielen oder der Versuch, die Korrespondenz von Topik und Subjekt bei der Konstitution eines Textes aufrechtzuerhalten.

Der vorgestellte Vermittlungsvorschlag versucht den Erwerbsproblemen durch Inputgestaltung und Bewusstmachung entgegenzuwirken. Die Zielkonstruktionen werden bei eindeutiger semantischer Situierung gehäuft angeboten und ihre gelenkte kommunikative Verwendung wird angestoßen. Die resultierenden Texte müssen in ihrer Hauptstruktur ein durchlaufendes Topik im Dativ erhalten. Ein notwendiger anschließender Schritt setzt für die Bewusstmachung eine Interpretationsaufgabe ein, um die Bedeutungen der Konstruktionen Unbeabsichtigter Verursachung und Nichtverhinderter Ereignisse, aber auch die des bloßen *Incommodi* genau zu differenzieren. Eine

explizitere Darstellung wie in Tabelle 1, die nach dem Muster traditioneller Grammatikarbeit metasprachlich verfährt und den strukturellen Vergleich mit den entsprechenden Transitiva und Intransitiva umfasst, ist je nach den Bedürfnissen der Lernenden möglich.

## Literatur

- De Bot, K., Ginsberg, R. & Krasmich, C. (Hrsg.). (1991). *Foreign language research in cross-cultural perspective*. Benjamins.
- Doughty, C. & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Hrsg.), *Focus on form in classroom second language acquisition* (S. 197–261). Cambridge University Press.
- Dowty, W. (1991). Thematic proto-roles and argument selection. *Language*, 67(3), 547–619.
- Dreyer, H. & Schmitt, R. (2009). *Lehr- und Übungsbuch der deutschen Grammatik – aktuell* (Neubearbeitung). Hueber.
- DWDS. *Das Digitale Wörterbuch der deutschen Sprache des 20. Jh.* (o. J.). Erarbeitet an der Berlin-Brandenburgischen Akademie der Wissenschaften; <https://www.dwds.de>.
- Eisenberg, P. (2020). *Grundriss der deutschen Grammatik. Der Satz* (5. Aufl.). Metzler.
- Fandrych, C. & Tallowitz, U. (2021). *Klipp und Klar – Neubearbeitung*. Klett.
- Fandrych, C. & Thurmair, M. (2021). *Grammatik im Fach Deutsch als Fremd- und Zweitsprache. Grundlagen und Vermittlung* (2. Aufl.). Erich Schmidt Verlag.
- Fausey, C. & Boroditsky, L. (2010). Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bull Rev.* 17, 644–650.
- Fausey, C. & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bull Rev.* 18, 150–157.
- Goldberg, A. (1995). *Constructions. A construction grammar approach to argument structure*. University of Chicago Press.
- Goldberg, A. (2006). *Constructions at work. The nature of generalization in language*. Oxford University Press.
- Hall, K. & Scheiner, B. (2020). *Deutsch – Übungsgrammatik für die Oberstufe – aktuell*. Hueber.
- Haspelmath, M. (1993). *A grammar of Lezgian*. Mouton de Gruyter.
- Helbig, G. & Buscha, J. (2013). *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Klett-Langenscheidt.
- Hens, G. (1997). Constructional semantics in German. The dative of inaction. *American Journal of Germanic Linguistics & Literatures*, 9(2), 191–219.
- Hoffmann, L. (2016). *Deutsche Grammatik. Grundlagen für Lehrerbildung, Schule, Deutsch als Zweitsprache und Deutsch als Fremdsprache* (3. Aufl.). Erich Schmidt Verlag.

- Hohmann, S. & Rohrmann, L. (2021). *Grammatik mal vier*. Klett.
- Jin, F. & Voß, U. (2013). *Grammatik aktiv A1–B1*. Cornelsen.
- Jin, F. & Voß, U. (2017). *Grammatik aktiv B2–C1*. Cornelsen.
- Kars, J. & Häussermann, U. (1988). *Grundgrammatik Deutsch*. Diesterweg.
- Kittilä, S. (2005). Remarks on involuntary agent constructions. *Word* 56(3), 381–419.
- Klein, W. & von Stutterheim, C. (1987). Quaestio und referentielle Bewegung in Erzählungen. *Linguistische Berichte* 109, 163–183.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. University of Chicago Press.
- Latour, B. (1988). *Mittelstufen-Grammatik für Deutsch als Fremdsprache*. Hueber.
- Long, M. (1991). Focus on form. A design feature in language teaching methodology. In K. De Bot, R. Ginsberg & C. Kramsch (Hrsg.), *Foreign language research in cross-cultural perspective* (S. 39–52). Benjamins.
- Lyons, J. (1977). *Semantics 1*. Cambridge University Press.
- Madlener, K. & Behrens, H. (2015). *Konstruktion(en) sprachlichen Wissens. Lernprozesse im Erst- und Zweitspracherwerb*. Arbeitspapier Universität Basel, Schweiz.
- Primus, B. (2012). *Semantische Rollen*. Winter.
- Reimann, M. (1996). *Grundstufen-Grammatik*. Hueber.
- Schulz, D. & Griesbach, H. (1995). *Grammatik der deutschen Sprache* (11. Aufl.). Hueber.
- Takahashi, M. (2017). „Affiziertheit“ und „unabsichtliche Kausierung“. Lesarten der Dativkonstruktionen bei Bewegungs- und Zustandsveränderungsverben. *Deutsche Sprache. Zeitschrift für Theorie, Praxis und Dokumentation* 45, 362–377.
- von Stutterheim, C. & Carroll, M. (2007). Durch die Grammatik fokussiert. *Zeitschrift für Literaturwissenschaft und Linguistik* 145, 35–60.
- Wegener, H. (1985). *Der Dativ im heutigen Deutsch*. Narr.
- Zifonun, G., Hoffmann, L. & Strecker, B. (1997). *Grammatik der deutschen Sprache*. 3 Bände. De Gruyter.



# Man darf nicht verallgemeinern: Überlegungen zur Angemessen- heit der (in)definiten NPn in generischen Kontexten

---

**Zusammenfassung** In diesem Beitrag werden (in)definite singularische und pluralische NPn unter der Perspektive der generischen Interpretation untersucht. Es wird gezeigt, dass das Generizitätspotenzial der NPn von semantisch-pragmatischen Restriktionen betroffen ist und bei jeder der NP-Formen unterschiedlich ausfällt. Dieser Artikel soll der Frage der ungleichmäßigen Distribution der NP-Formen in Hinblick auf die Artikelsemantik nachgehen. Im Rahmen der Einzigkeitstheorie zur Bedeutung des definiten Artikels soll der logische Zusammenhang zwischen Definitheit und Gattungsreferenz aufgedeckt werden. Die darauf aufbauende Annahme zur Semantik der Indefiniten soll ihrerseits begründen, warum Gattungen nicht indefinit sein können. Die theoretischen Schlüsse werden mit einer Reihe von Beispielen, die semantische und pragmatische Kontraste zwischen den NP-Formen hervorheben, belegt.

**Schlüsselwörter** Generizität; NP-Formen; Gattungsreferenz; Gattungsprädikate; Charakterisierungsprädikate; abstrakte Gattung; taxonomisch; Artikelsemantik; Einzigkeit; potenzielle Referenten

## 1 Einleitung

Krifka et al. (1995) unterscheiden zwei Generizitätstypen<sup>1</sup>: die charakterisierende Aussage (engl. *characterizing sentence*) und die Gattungsreferenz (engl. *kind reference*). Die charakterisierende Aussage stellt eine Generalisierung dar, die eine bestimmte Regularität ausdrückt. Beispielsweise

---

1 Für alternative Klassifizierungen siehe Chur (1993); Grimm (1986); Radden (2009).

charakterisiert die Aussage (1) einen Laserdrucker aufgrund der von partikulären Fakten abstrahierten Regularität, für den Fotodruck nicht geeignet zu sein.

(1) *Ein Laserdrucker* ist für den Ausdruck von Fotos nicht geeignet.

Das Kennzeichen charakterisierender Aussagen sind Charakterisierungsprädikate (engl. *characterizing predications*). Charakterisierungsprädikate können sehr vielfältig sein. Ihr gemeinsames Merkmal ist der Ausdruck einer, so Krifka et al. (1995, S. 19), „essenziellen Eigenschaft“ der Entität, über die eine charakterisierende Aussage gemacht wird. Nach Behrens (2005) hingegen muss das Charakterisierungsprädikat nicht unbedingt eine essenzielle, aber eine prototypische Eigenschaft darstellen, denn charakterisierende Aussagen sind gegenüber Ausnahmen tolerant. So muss das Prädikat *ist für den Ausdruck von Fotos nicht geeignet* nicht zwangsläufig auf alle Laserdrucker zutreffen. Behrens (ebd.) weist darauf hin, dass Ausnahmen als solche nur dann erkannt werden können, wenn sie individualisiert sind. Beispielsweise wenn auf die charakterisierende Aussage (1) die Behauptung über einen spezifischen Laserdrucker folgt: „Ich habe einen Laserdrucker auch für den Ausdruck von Fotos gefunden“, wird ihr positiver Wahrheitswert nicht im Widerspruch zu (1) stehen.

Der andere Generizitätstyp, die Gattungsreferenz, ergibt sich als Generalisierung auf der NP-Ebene. Das bedeutet, dass eine NP auf eine von partikulären Objekten abstrahierte Gattung referiert. Als Beispiel einer gattungsreferierenden NP ist die NP in (2) zu betrachten:

(2) *Der Laserdrucker* wurde im Jahr 1969 erfunden.

Gerstner-Link (1995) weist darauf hin, dass keine Sprache der Welt ein Kodierungselement besitzt, das ausschließlich als Generizitätsmarker einer NP fungieren könnte. So kann die definite NP *der Laserdrucker* abhängig vom Kontext als partikulär, d. h. objektreferierend, oder als gattungsreferierend interpretiert werden. Die Generizität wird demnach nicht als inhärente NP-Eigenschaft gesehen, sondern als ihre Lesart. In diesem Beitrag wird der Frage nachgegangen, wie die generische Lesart mit der Semantik (in)definiten singularischer und pluralischer NPn korreliert. Nach einer groben Skizzierung von semantisch-pragmatischen Restriktionen für das generische Potenzial der vier NP-Formen wird der Versuch unternommen, ihre referentielle Bedeutung unter der Perspektive der Artikelsemantik festzulegen und mit Hinblick darauf die Einschränkungen zu erklären.

## 2 NP-Einschränkungen bei Gattungsprädikaten

Auf die generische Lesart einer NP weisen die sog. Gattungsprädikate (engl. *kind predications*) hin. Nach Krifka et al. (1995) können Gattungsprädikate lexikalisch oder abgeleitet sein. Zu den lexikalischen Gattungsprädikaten gehören solche Verben wie *aussterben*, *sich sammeln*, *erfinden* und prädikative Adjektive wie *üblich sein*, *verbreitet sein*, *selten sein* etc. Diese beziehen sich direkt auf eine Gattung und können folglich nur in einem Kontext mit einer gattungsreferierenden NP gebraucht werden: Es kann nämlich nur eine Gattung aussterben oder verbreitet sein, aber kein Exemplar dieser Gattung. Die Kombinierbarkeit der NP-Formen mit lexikalischen Gattungsprädikaten wird daher als diagnostischer Test benutzt, um die Formen einer gattungsreferierenden NP zu bestimmen (vgl. ebd., S. 79), z. B.:

- (3) *Der Löwe* ist in Afrika verbreitet.
- (4) *!Ein Löwe* ist in Afrika verbreitet.<sup>2</sup>
- (5) *Die Löwen* sind in Afrika verbreitet.
- (6) *Löwen* sind in Afrika verbreitet.

Beim ersten Blick auf die Beispielgruppe fällt auf, dass der indefinite Singular an diesem Test scheitert. Dies belegt die Annahme von Krifka et al. (1995), dass indefinite singularische NPn nur auf Objekte referieren können. Auch bei der angemessenen Verwendung in charakterisierenden Aussagen wie (1) sei der indefinite Singular nur scheinbar generisch: „The locus of genericity is not in the indefinite singular NP, but rather in the sentence itself; and therefore this type of genericity is outside the nominal system” (Krifka et al., ebd., S. 14). Im Gegensatz zum indefiniten Singular können alle anderen NP-Formen mit Gattungsprädikaten uneingeschränkt auftreten, was ihnen in generischen Kontexten<sup>3</sup> den Status einer gattungsreferierenden NP verleiht. Wenn man jedoch den Fokus auf die Kombinierbarkeit der definiten NPn und indefiniten Plurale mit sog. abgeleiteten Gattungsprädikaten richtet, sieht man eine ungleichmäßige Distribution der NP-Formen. Die ableiteten Gattungsprädikate sind Prädikate, die zwar in partikulären Kontexten mit einer objektreferierenden NP auftreten können, aber mit einer gattungsreferierenden NP als Gattungsprädikate interpretiert werden. Je nach der Semantik dieser Prädikate unterscheiden Krifka et al. (ebd.) sieben unterschiedliche Interpretationen, die Sätze erhalten können. Im Folgenden werden die

---

2 Mit dem Ausrufezeichen wird markiert, dass der Satz weder Gattungsreferenz enthält noch als charakterisierende Aussage auftritt.

3 Als generische Kontexte werden hier Sätze mit Gattungs- bzw. Charakterisierungsprädikaten bezeichnet.

Interpretationen beschrieben und hinsichtlich ihrer Kombinationsmöglichkeiten mit den vier NP-Formen analysiert.

Als erste ist die *collective property*-Interpretation zu betrachten. In diesem Fall bezieht sich das Prädikat auf die Gesamtmenge von Objekten, die zu einer bestimmten Gattung gehören. Wie die folgende Beispielgruppe veranschaulicht, können nicht alle NP-Formen eine angemessene Interpretation evozieren:

- (7) ?*Der deutsche Käufer*<sup>4</sup> hat letztes Jahr 11000 BMWs gekauft.<sup>5</sup>
- (8) *Die deutschen Käufer* haben letztes Jahr 11000 BMWs gekauft.
- (9) !*Ein deutscher Käufer* hat letztes Jahr 11000 BMWs gekauft.
- (10) !*Deutsche Käufer* haben letztes Jahr 11000 BMWs gekauft.

In der Arbeit von Molnár (2014) wird behauptet (allerdings ohne anschließende Argumentation), dass in den Sätzen mit einer *collective property*-Interpretation der indefinite Singular ausgeschlossen und dass der definite nicht fraglos sei. Die angemessenste Form stelle die definite pluralische NP dar, wobei auch der indefinite Plural akzeptabel sei (vgl. Molnár, ebd., S. 147). Die Beobachtungen scheinen bis auf (10) plausibel zu sein. Der Satz (10) wird eher so verstanden, dass einzelne deutsche Käufer 11000 BMWs gekauft hätten, was nicht realistisch ist. In der einschlägigen Literatur finden sich Hinweise darauf, dass der indefinite Plural im Gegensatz zum definiten nicht die kollektive, sondern die distributive Lesart erzeugt (vgl. Laca, 1992; Kefer, 2011; Schaden, 2013). Dieser Befund dient somit als Erläuterung für unsere Intuition, dass die indefinite pluralische NP zur *collective property*-Interpretation der Sätze nicht beitragen kann.

Wenn der Durchschnitt der Gattung präzisiert wird, geht es um die *average property*-Interpretation, vgl.:

- (11) *Die deutschen Teenager* sehen täglich 6 Stunden lang fern.

Kontrastiert man das Beispiel (11) mit (8), lassen sich keine syntaktischen Unterschiede feststellen, auf die die Differenzen in der Interpretation zwischen den Sätzen zurückgeführt werden könnten. Die unterschiedlichen Interpretationen entstammen dem Faktenwissen (vgl. Krifka et al., 1995, S. 80). Tatsächlich ist es sehr unwahrscheinlich, dass die Gesamtheit von deutschen Teenagern nur sechs Stunden lang fernsieht, wie das im Fall der

---

4 Mit dem Fragezeichen wird markiert, dass eine NP-Form unnatürlich klingt bzw. unangemessen scheint.

5 (8) ist das Originalbeispiel von Krifka et al. (1995, S. 78), Übersetzung von Molnár (2014, S. 147); (7), (9), (10) sind Modifizierungen von Molnár (ebd.).



*collective property*-Interpretation wäre. Logischerweise ist hier die Rede von einem durchschnittlichen deutschen Teenager. In Sätzen mit der *average property*-Interpretation können außer dem definiten Plural auch der indefinite Plural und der definite Singular vorkommen (vgl. Molnár, 2014, S. 147):

- (12) *Deutsche Teenager* sehen täglich 6 Stunden lang fern.<sup>6</sup>
- (13) *Der deutsche Teenager* sieht täglich 6 Stunden lang fern.
- (14) *?Ein deutscher Teenager* sieht täglich 6 Stunden lang fern.

Im Gegensatz zu den Definiten ist der indefinite Singular mit der *average property*-Interpretation nicht akzeptabel. Sein Auftreten wird jedoch durch das Einfügen eines nominalen Adjektivs wie *typisch*, *durchschnittlich*, *gewöhnlich* etc. ermöglicht (vgl. Krifka et al., 1995, S. 80):

- (15) *Ein typischer deutscher Teenager* sieht täglich 6 Stunden lang fern.<sup>7</sup>

Die Sätze, in denen einer Gattung eine prototypische Eigenschaft zugewiesen wird, erhalten eine sog. *characterizing property*-Interpretation. Die charakteristische Eigenschaft drücken Charakterisierungsprädikate aus. Da sie den zweiten Generizitätstyp, die charakterisierende Aussage, kennzeichnen, sprechen Krifka et al. (ebd.) in diesem Fall von der Kombination der Gattungsreferenz mit der charakterisierenden Aussage. Die Beispiele (16)–(18) zeigen, dass die gattungsreferierende NP mit Charakterisierungsprädikaten alle ihre möglichen Formen annehmen kann:

- (16) *Der Braunbär* ist gefährlich.
- (17) *Die Braunbären* sind gefährlich.
- (18) *Braunbären* sind gefährlich.

Wenn eine indefinite singularische NP mit Charakterisierungsprädikaten vorkommt, tritt eine charakterisierende Aussage eigenständig auf:

- (19) *Ein Braunbär* ist gefährlich.

Aussagen mit *characterizing property*-Interpretation stehen solchen mit *distinguishing property*-Interpretation nahe. Auf den ersten Blick lassen sich die Interpretationen kaum unterscheiden, denn beide drücken eine Eigenschaft

---

6 (12) ist das Originalbeispiel von Krifka et al. (1995, S. 78), Übersetzung von Molnár (2014, S. 146); (11), (13), (14) sind Modifizierungen von Molnár (ebd.).

7 Beispiel von Krifka et al. (1995, S. 80), Übersetzung von Molnár (2014, S. 147).

aus, die mit einer Gattung assoziiert wird. Der Unterschied wird sichtbar, wenn man beide Typen von Aussagen unter der Perspektive des Wahrheitsgehalts vergleicht:

<p>(20) <i>characterizing</i> property-Interpretation</p> <p>a) <i>Die Kartoffel</i> enthält Vitamin C.<sup>8</sup></p> <p>b) <i>Die Kartoffeln</i> enthalten Vitamin C.</p> <p>c) <i>Kartoffeln</i> enthalten Vitamin C.</p> <p>d) <i>Eine Kartoffel</i> enthält Vitamin C.</p>	<p>(21) <i>distinguishing</i> property-Interpretation</p> <p>a) <i>Der Deutsche</i> ist im Job effizient.</p> <p>b) <i>Die Deutschen</i> sind im Job effizient.</p> <p>c) <i>Deutsche</i> sind im Job effizient.</p> <p>d) <i>?Ein Deutscher</i> ist im Job effizient.</p>
--	--

Alle vier Aussagen in (20) haben den gleichen Wahrheitswert, d. h. sowohl für die Gattung „Kartoffel“ (20a–c) als auch für ein arbiträres Exemplar dieser Gattung (20d) ist der Gehalt vom Vitamin C wahr. Betrachtet man die Beispiele in (21), dann lässt sich feststellen, dass hier der Wahrheitswert unterschiedlich ausfällt. Während die Aussagen (21a–c) implizieren, dass die Deutschen für ihre Effizienz im Job bekannt sind, setzt (21d) voraus, dass ein arbiträrer Vertreter der deutschen Nationalität im Job effizient ist. Die Sätze mit den gattungsreferierenden NPn haben daher mehr Veranlagung, als wahr interpretiert zu werden, im Vergleich zu der Aussage mit der objektreferierenden NP, die nach Krifka et al. „a stronger claim“ (1995, S. 81) aufstellt. Die Beispiele (21a–c) sind als Fälle der *distinguishing property*-Interpretation zu behandeln, weil die hier ausgedrückte Eigenschaft Entitäten unterscheidet, die zur gleichen Kategorie (wie z. B. Nationalitäten) gehören.

Wenn das Prädikat von einem Objekt als Repräsentant einer Gattung auf die Gattung projiziert wird, hat der Satz eine *representative object*-Interpretation. Vgl.:

(22) Hier wurde in diesem Sommer *der schwarze Schwan* gesichtet.<sup>9</sup>

In der Situation, die mit (22) beschrieben wird, mag einer oder mehrere schwarze Schwäne gesichtet worden sein; aber im Satz handelt es sich nicht um die Schwäne, sondern um die abstrakte Gattung „Schwarzer Schwan“, die

8 (20a) ist das Originalbeispiel von Krifka et al. (1995, S. 78), Übersetzung von Molnár (2014, S. 148); (20b–d) sind Modifizierungen von Molnár (ebd.).

9 Beispiele von Gerstner-Link (1995, S. 91).

durch die konkreten Schwäne repräsentiert wird. Unter den NP-Formen in den Sätzen mit der *representative object*-Interpretation scheint eine definite singularische NP am angemessensten zu sein. Der definite Plural ist nach Gerstner-Link „nicht fraglos glatt“ (1995, S. 91), weil die Form zur Summenlesart tendiere, die auf konkrete Individuen ziele. Auf konkrete Individuen würden sich der indefinite Singular und indefinite Plural beziehen, deshalb kann auch hier keine Rede von Generizität sein. Aus diesem Grunde werden indefinite NPn aus der *representative object*-Interpretation ausgeschlossen (ebd.). Vgl.:

- (23) ?Hier wurden in diesem Sommer *die schwarzen Schwäne* gesichtet.
- (24) !Hier wurde in diesem Sommer *ein schwarzer Schwan* gesichtet.
- (25) !Hier wurden in diesem Sommer *schwarze Schwäne* gesichtet.

Von der *representative object*-Interpretation ist die *avant-garde*-Interpretation zu unterscheiden. Die Besonderheit solcher Sätze besteht in der Übertragung einer „historisch-akzidentellen Eigenschaft“ (Gerstner-Link, 1995, S. 90) von einer bestimmten Avantgarde (Objekt bzw. Gruppe von Objekten) auf die entsprechende Gattung. So kann der Verdienst der ersten Everest-Besteiger E. Hillary und T. Norgay aufgrund seiner großen Relevanz auf die ganze Menschengattung übertragen werden. Die Beispiele (26)–(29) zeigen, dass die *avant-garde*-Interpretation nur die Sätze mit den definiten NPn erhalten. Die indefiniten NPn beziehen sich genau auf die Menschen, die den Everest bestiegen haben.

- (26) 1953 bestieg *der Mensch* den höchsten Berg der Erde.<sup>10</sup>
- (27) 1953 bestiegen *die Menschen* den höchsten Berg der Erde.
- (28) !1953 bestieg *ein Mensch* den höchsten Berg der Erde.
- (29) !1953 bestiegen *Menschen* den höchsten Berg der Erde.

Eine Aussage nimmt eine *internal comparison*-Interpretation an, wenn die Gattung mit sich selbst innerhalb einer Dimension verglichen wird. Beispiele für eine solche Interpretation stellen die Sätze mit den definiten NPn in (30) und (31) dar. Beide Aussagen bedeuten: Wenn man nach Norden fährt, wird die Gattung „Wolf“ durch größere Exemplare vertreten.

- (30) *Der Wolf* wird größer, wenn wir Richtung Norden fahren.<sup>11</sup>
- (31) *Die Wölfe* werden größer, wenn wir Richtung Norden fahren.

---

10 Beispiele von Gerstner-Link (1995, S. 91).

11 (31) ist das Originalbeispiel von Krifka et al. (1995, S. 78), Übersetzung von Molnár (2014, S. 150); (30), (32), (33) sind Modifizierungen von Molnár (ebd.).

Die Sätze mit den indefiniten NPn hingegen können keine *internal comparison*-Interpretation aufweisen (vgl. von Frieling, 2002; Molnár, 2014):

(32) *!Ein Wolf* wird größer, wenn wir Richtung Norden fahren.

(33) *!Wölfe* werden größer, wenn wir Richtung Norden fahren.

Unter den Beispielen (32) und (33) sei „nur eine unsinnige Aussage zu verstehen, dass die spezifischen<sup>12</sup> Wölfe, die man auf einer Reise mit sich führt, immer mehr wachsen desto weiter man nach Norden gelangt“ (von Frieling, 2002, S. 41).

In Tabelle 1 werden die Ergebnisse der durchgeführten Beurteilung der NP-Formen hinsichtlich ihrer Angemessenheit in Kontexten mit lexikalischen und abgeleiteten Gattungsprädikaten zusammengefasst.

Tabelle 1. NP-Formen in Kontexten mit Gattungsprädikaten  
(eigene Darstellung in Anlehnung an Molnár, 2014)

Prädikatstyp	Interpretation	Definite NP		Indefinite NP	
		SG.	PL.	SG.	PL.
Lexikalisches Gattungsprädikat	kind predication	+	+	-	+
Abgeleitetes Gattungsprädikat	avant-garde	+	+	-	-
	average property	+	+	?	+
	chrakterizing property	+	+	+	+
	collective property	?	+	-	-
	distinguishing property	+	+	?	+
	internal comparison	+	+	-	-
	respresentative object	+	?	-	-

Aus der Tabelle wird ersichtlich, dass jede der Erscheinungsformen einer gattungsreferierenden NP von Restriktionen der abgeleiteten Gattungsprädikate betroffen ist. Vor dem Hintergrund der unterschiedlichen Distributionen der gattungsreferierenden NP-Formen ergibt sich die Frage nach den Gründen für solche Restriktionen. Man kann annehmen, dass die drei NP-Formen eine

12 Unter „spezifisch“ wird hier höchstwahrscheinlich „partikulär“ verstanden, sodass das Merkmal in der Opposition zu „generisch“ steht, und nicht zur unspezifischen Referenz.

unterschiedliche referentielle Semantik tragen, die nicht zu allen Gattungsprädikaten ‚passt‘. Im aktuellen Zusammenhang wird davon ausgegangen, dass die NP-Einschränkungen auf zwei Faktoren der NP-Semantik zurückzuführen sind: Artikel und Numerus. Mit der nachfolgenden Analyse soll diese Annahme belegt werden.

### 3 Zusammenhang zwischen Definitheit und Generizität

#### 3.1 Semantik des definiten Artikels

Einen elaborierten theoretischen Ansatz zur Beschreibung der Artikelsemantik bietet Hawkins (1991) an. Hawkins greift die Annahme von Russell (1905) auf, der zufolge definite Ausdrücke mit der Einzigkeit ihres Referenten einhergehen, und erweitert sie unter Einbeziehung von pragmatischen Aspekten:

At the one extreme uniqueness may hold within the universe of discourse as a whole. Alternatively (and much more frequently) it may be satisfied only within a proper subset of the universe, in which case the hearer must be able to infer the pragmatic parameters intended by the speaker, in form of some P-Set. (Hawkins, 1991, S. 412)

Laut Hawkins (ebd.) soll das P-Set, d. h. das pragmatische Set, wechselseitig manifest sein. Den Begriff wechselseitiger Manifestiertheit entnimmt Hawkins aus der Relevanztheorie von Sperber und Wilson (1986), der zufolge eine Tatsache nur dann manifest ist, wenn von ihr bei einem Individuum eine zutreffende oder wahrscheinlich zutreffende Repräsentation evoziert wird (ebd., S. 39). Was den Begriff des P-Sets betrifft, so geht Hawkins nicht detailliert auf ihn ein. Das Konzept ist jedoch intuitiv klar: P-Sets sind höchstwahrscheinlich das, was Sperber und Wilson (ebd.) mit dem Terminus „kognitive Umgebung“ bezeichnen. Die kognitive Umgebung umschreiben die Wissenschaftler wie folgt: „a cognitive environment of an individual is a set of facts that are manifest to him“ (ebd.). Als wechselseitig manifestes P-Set (nach Sperber und Wilson wechselseitige kognitive Umgebung) wird also „any shared cognitive environment in which it is manifest which people share it ... [and in which] ... every manifest assumption is what we will call mutually manifest“ (ebd., S. 41–42) verstanden. Hawkins (1978) weist darauf hin, dass ein P-Set auf textueller, assoziativer sowie situativer Grundlage entstehen kann, und unterscheidet vier Typen von Sets:

- Set des bisherigen Diskurses: „Ich war bei einer Hochzeit. *Die Hochzeit* war in Paris.“
- assoziatives Set: „Ich war bei einer Hochzeit. *Die Braut* sah wunderschön aus.“
- unmittelbar situatives Set: „Könnten Sie mir bitte *den Pfeffer* geben?“
- größeres situatives Set: „Waren Sie *in der Staatsoper*?“

Eine definite NP wird verwendet, wenn sich in einem von diesen P-Sets ein einziger potenzieller Referent dieser NP befindet. So ist z. B. *die Hochzeit* definit, weil im P-Set des bisherigen Diskurses nur ein potenzieller Referent der NP vorhanden ist.

Die Einzigkeitsbedingung des definiten Artikels lässt sich auch auf pluralische NPn und NPn mit Massennomen erweitern. Dafür sei der Referent einer definiten NP als einziges maximales Set von Entitäten aufzufassen (vgl. Hawkins, 1991, S. 410). Das einzige maximale Set bedeutet, dass zu dem Set alle Entitäten gehören, die unter die jeweilige Beschreibung fallen und da es keine Entität außerhalb des Sets gibt, ist das Set einzig. Insofern geht es um ein Set als kollektives Ganzes, d. h. als komplexes Objekt, dessen Grenzen mit dem letzten Element enden, das unter das entsprechende Konzept fällt. Empirisch lässt sich diese Auffassung mit der Änderung der Wahrheitsbedingungen beim Ersatz des definiten Artikels durch den *all*-Quantor belegen. Vgl.:

(34) *Die Studenten* an meiner Uni sind faul.

(35) *Alle Studenten* an meiner Uni sind faul.

Im Unterschied zu (34) ist die Aussage (35) auch wahr, wenn nicht alle Studenten an der Uni des Sprechers faul sind. Das bedeutet, dass die definite Referenz keine Quantifikation über die Studenten als Elemente des entsprechenden Sets voraussetzt (vgl. ebd.). Die Aussage bekommt den negativen Wahrheitswert nur dann, wenn das Prädikat keine charakteristische Eigenschaft für das Kollektiv „Studenten an meiner Uni“ darstellt.

Neben pragmatischer Einzigkeit unterscheidet Hawkins semantische und ontologische Einzigkeit. Auf die zwei Einzigkeitstypen weisen definite Verwendungen mit Modifikatoren hin, die einen sog. „inherent uniqueness claim“ (ebd., S. 420) enthalten. Die Sätze (36)–(37) illustrieren den Fall der Referenteneinzigkeit, die sich aus der Semantik des NP-Modifikators ergibt:

(36) Lisa und Anna haben *den gleichen Laptop*.

(37) *Das fleißigste Tier* ist die Honigbiene.

Der attributive Modifikator *gleich* setzt voraus, dass es nur eine Entität geben kann, die diese Eigenschaft trägt. Laut Hawkins (ebd.) lässt sich das Konzept

des Gleichseins als einzigartiger Typ abbilden, zu dem zwei oder mehrere Tokens gehören. Der Satz (36) bedeutet somit nicht, dass Lisa und Anna gemeinsam einen Laptop besitzen, sondern dass sie denselben Typ von Laptop haben. In (37) ist der einzigkeitsimplizierende Modifikator durch den Superlativ vertreten. Der Superlativ signalisiert die Einzigkeit der Entität, die die superlative Eigenschaft trägt, und erfordert deshalb eine definite NP für deren Kennzeichnung.

In den Beispielen (38)–(39) ist die Verwendung einer definiten NP durch ontologische Einzigkeit des Referenten motiviert:

(38) Peter hat ein Haus *auf dem Gipfel des Berges*.

(39) *Die Mutter von Anna* arbeitet im Krankenhaus.

Die Annahmen über die ontologische Einzigkeit beruhen auf unserem Allgemeinwissen über den Aufbau der realen Welt. So ergibt sich die Einzigkeit der definiten Referenten in (38)–(39) aufgrund der allgemeinbekannten Tatsache, dass ein Berg i. d. R. nur einen Gipfel bzw. eine Person nur eine Mutter hat. Zu den ontologisch einzigen Referenten gehören außerdem Referenten von NPn mit sog. Unika: der Mond, der Himmel, der Südpol etc. Während sich die Einzigkeit der Referenten von (38)–(39) nur relativ zum Referenten des entsprechenden Modifikators manifestiert (Berg bzw. Anna), wird die Einzigkeit von den angeführten Unika-Referenten relativ zur ganzen Erde angenommen. Da die Erde als Anker für die Einzigkeit der Unika uns als Erdenbewohnern meistens kontextunabhängig wechselseitig manifest ist, erscheint eine explizite Einschränkung der Referenteneinzigkeit wie *der Himmel über der Erde* oder *der Südpol der Erde* überflüssig.

### 3.2 Warum ist die Gattung definit?

Als Hauptthese des Einzigkeitsansatzes von Hawkins (1991) ist hervorzuheben, dass der Referent einer definiten NP einzig sein muss und dass seine Einzigkeit pragmatisch, semantisch oder ontologisch bedingt sein kann. Nun ist der Typ der Einzigkeit festzustellen, der den Gattungen als definiten Referenten zugrunde liegt. Manche Linguisten (z. B. Bisle-Müller, 1991; Behrens, 2005; Mumm, 1995; Prince, 1992) gehen davon aus, dass Gattungen aufgrund des Allgemeinwissens im Diskurs permanent vorhanden sind und daher für Hörer immer etwas Bekanntes darstellen. Diese Annahme scheitert jedoch an der Erläuterung von definiten NPn, die für Gattungen stehen, bei welchen Unbekanntheit unterstellt wird, vgl.:

(40) Zu den seltensten Gattungen gehört *der Vaquita*. Das ist ein kleiner Schweinswal, der im Golf von Mexiko beheimatet ist.

- (41) An der Spitze der Republik Venedig stand *der Doge*. Das war das Staatsoberhaupt, das uneingeschränkte Macht besaß.<sup>13</sup>

Die an die definite NP anschließende Erklärung, was unter *Vaquita* bzw. *Doge* zu verstehen ist, deutet darauf hin, dass die Gattungsreferenz ohne Link zum bisherigen Diskurs erfolgt (vgl. Vater, 1997). Also können Gattungen nicht pragmatisch einzig sein.

In Christophersen (1939) finden sich Hinweise darauf, dass Definitheit gattungsreferierender NPn mit ontologischer Einzigkeit des Referenten verbunden sein kann. Laut Christophersen ist eine definite NP nur dann gattungsreferierend, wenn sie mit Unika auftritt. So referiere die definite NP *der Mond* auf das Objekt „Mond“ und, da das Objekt einzigartig ist, auf die entsprechende Gattung. Das Nomen in der NP *der Löwe* ist zwar kein Unikum, wenn es aber eine Gattung bezeichnen soll, wird es als solches verstanden. In diesem Fall wird die Gattung als einziges Objekt „Löwe“ repräsentiert: „When we say ‚The lion is the king of beasts‘, we imagine for a moment that there is only one lion, which is in itself the whole species, and similarly with the other beasts of which the lion is the king“ (Christophersen, 1939, S. 77). Christophersen erläutert allerdings nicht, auf welcher Grundlage das Rezipieren der Gattung, die in der Realität durch mehrere Exemplare vertreten wird, durch ihren hypothetisch einzigen Repräsentanten erfolgt. Neben der fehlenden Argumentation bei der Einordnung der gattungsreferierenden NP mit Gattungsnamen scheint selbst die Grundannahme, der zufolge Unika gattungsreferierend sind, problematisch zu sein. Wenn Christophersen eine Gattung als die Gesamtheit ihrer Exemplare versteht, kann seine Folgerung, nach welcher das Mond-Objekt die ganze Mondgattung ist, nachvollzogen werden. Das bedeutet aber nicht, dass die NP *der Mond* gattungsreferierend ist, sondern lediglich, dass all die durch das Prädikat dem Mond-Objekt zugeschriebenen Eigenschaften auch auf die ganze Mondgattung zutreffen, sodass man davon ausgehen kann, dass eine Aussage über ein einzigartiges Exemplar indirekt eine Aussage über die ganze Gattung darstellt.

Die Gattung kann sich aber auch als abstrakte Entität manifestieren, etwa als „second order individual“ (Lyons, 1999, S. 179). Unter ebendieser Auffassung der Gattung ist die NP *der Löwe* in Christophersens Aussage zu interpretieren. Der definite Singular *der Löwe* ist gattungsreferierend, weil er für eine abstrakte Gattung steht, und definit, weil eine abstrakte Gattung logisch bedingt einzig ist. Es kann keine mehreren gleichen Gattungen geben: Alles, was als Gattung „Löwe“ bezeichnet wird, ist ein und dieselbe Gattung „Löwe“. Aus diesem Grund erfordert die Gattung als Diskursreferent eine definite Kennzeichnung. Krifka et al. (1995) weisen darauf hin, dass die Gattungen,

13 Beispiel in Anlehnung an Vater (1997, S. 222).



auf die referiert wird, nur sog. gut etablierte Gattungen (engl. *well-established kinds*) sein können (vgl. Krifka et al., 1995, S. 11). Zur Unterscheidung der gut etablierten Gattungen von den nicht-gut etablierten führen die Linguisten folgende Beispiele an:

- (42) *The Coke bottle has a narrow neck.*  
 (43) *?The green bottle has a narrow neck.*

Es gibt die in der westlichen Kultur etablierte Gattung „Colaflasche“, aber keine gut etablierte Gattung „Grüne Flasche“. Das ist der Grund, weshalb der Ausdruck (43) als inakzeptabel gesehen wird. Für den aktuellen Zusammenhang ist es wichtig zu betonen, dass die illustrierten Einschränkungen ausschließlich pragmatische Unangemessenheit aufweisen. Semantisch gesehen ist die Verwendung des definiten Singulars nicht nur angemessen, sondern erforderlich, solange es um die abstrakte Gattung „Grüne Flasche“ geht. Die pragmatische Unangemessenheit solcher definiten Verwendung hängt wohl mit der Auffassung einer heterogenen Kategorie wie grünen Flaschen als abstrakte Gattung „Grüne Flasche“ zusammen (vgl. Radden, 2009). Die Abstraktion einer Gattung scheint einen hohen Grad an Homogenität der entsprechenden Gruppe zu verlangen, sodass in der mentalen Repräsentation bei Sprechern ein möglichst konkretes Bild bei der Erwähnung dieser Gattung evoziert werden kann. Solch ein Bild stellt, wie bereits bei Christophersen angemerkt wurde, das prototypische Exemplar der Gattung dar, und ohne es kann der Inhalt der Aussage nicht in vollem Umfang nachvollzogen werden. Dies ist vergleichbar mit dem Fall, wenn der Sprecher ein Wort benutzt, dessen Bedeutung weder ihm noch dem Hörer gut bekannt ist. Das folgende Satzpaar stellt ein weiteres Beispiel dar, aus dem die Rolle der Homogenität in der Abstraktion einer Gattung ersichtlich wird:

- (44) *?Der Baum ist hoch.*  
 (45) *Die Weichselkirsche wird bis zu 10 Meter hoch.*

Die Kategorie „Baum“ kann wegen ihrer Heterogenität kein Bild vom prototypischen Baumexemplar entstehen lassen und erscheint somit als abstrakte Gattung unplausibel. Demgegenüber stellt die Kategorie „Weichselkirsche“ aufgrund ihrer spezifischen Charakteristika eine hochhomogene Kategorie dar und eignet sich somit zur Abstraktion der entsprechenden Gattung gut.

Anschließend gehen wir der Frage zur Bedeutung des generischen definiten Plurals nach. Der definite Plural scheint für die Form der Gattungsreferenz, in der die Gattung die Gesamtheit ihrer Exemplare darstellt, verantwortlich zu sein. So referiert nach Radden (2009) der generische definite Plural auf das

ganze Set von Entitäten, das sich durch den pragmatischen Kontext abgrenzen lässt. Die Beispiele (46) und (47) veranschaulichen diese Annahme:

(46) *Die Amerikaner* haben den Mond 1969 betreten.

(47) *Die Amerikaner* sind tolerant.

Radden (2009) geht davon aus, dass die Aussage (46) alle amerikanischen Astronauten sowie NASA-Mitarbeiter betrifft, und weist deshalb darauf hin, dass das Set „Amerikaner“ innerhalb der Domäne Nationalität, Wohnsitz und Aeronautik abgegrenzt wird. Das Set „Amerikaner“ in (47) hingegen impliziert die USA-Bewohner mit amerikanischer Herkunft. Also ergebe es sich in der Domäne der Nationalität und des Wohnsitzes und sei somit erheblich größer als in (46). Die Interpretation von Radden enthält zwei Aspekte, denen im aktuellen Zusammenhang nicht zugestimmt werden kann: die pragmatische Abgrenzung des Sets sowie die Quantifikation über seine Elemente. Wenn es sich in den beiden Beispielen um die Referenz auf alle Entitäten im pragmatisch abgegrenzten Set handelt, dann kann man sich fragen, ob diese Aussagen überhaupt eine generische Interpretation haben (vgl. Hawkins, 1978): Bei partikulärer Referenz bezieht man sich nämlich auch auf ein pragmatisch abgegrenztes Set (siehe Abschnitt 3.1). Allerdings lassen sich bereits aus einer Umformulierung der Beispiele klare Evidenzen für generische, ergo pragmatisch uneingeschränkte, Referenz gewinnen, vgl.:

(48) *Ihr* habt den Mond 1969 betreten.

Wenn man sich vorstellt, dass (48) im Gespräch mit einem Amerikaner auftaucht, der in die Mondlandung keineswegs involviert war, erweist sich die NP-Verwendung trotzdem als pragmatisch angemessen. Diese Beobachtung deutet darauf hin, dass der NP-Skopos im vorliegenden Kontext nicht mit den Grenzen einer bestimmten Gruppe von Amerikanern endet, sondern sich über alle Amerikaner erstreckt. Das bedeutet, dass die NP als gattungsreferierend zu betrachten ist. Der Grund, aus welchem die Aussage ihren positiven Wahrheitswert im gegebenen Kontext nicht verliert, liegt in der kollektiven Lesart des definiten Plurals (siehe Abschnitt 3.1). Mit der NP *ihr*, mit der die Amerikaner impliziert werden, referiert der Sprecher nicht auf alle Amerikaner, sondern auf „Amerikaner“ als Nation oder allgemeiner gesagt als Kollektiv, zu dem alle Amerikaner einschließlich des amerikanischen Gesprächspartners gehören. Dass auch in (47) mit dem definiten Plural keine Quantifikation über alle Elemente der Gattung vorausgesetzt wird, veranschaulicht seine Umformulierung in (49):

(49) *Ihr* seid tolerant.

Selbst wenn der Sprecher seinen amerikanischen Gesprächspartner als intolerant einschätzt, ist die Aussage wahr, solange das Prädikat auf das Kollektiv „Amerikaner“ zutrifft. Unter dieselben Wahrheitsbedingungen fallen (50) und (51):

- (50) *Die Italiener* machen die besten Möbel.  
 (51) *Die Tiger* setzen sich aus neun Unterarten zusammen.<sup>14</sup>

Im Satz (50) handelt es sich offensichtlich nicht um alle Italiener, sondern um die Italiener als Kollektiv aufgrund nationaler Zugehörigkeit. Noch deutlicher wird die kollektive Referenz in (51). Neben der NP signalisiert das Prädikat *setzen sich aus neun Unterarten zusammen*, dass es nicht um alle Exemplare geht, sondern um „Tiger“ als Kollektiv. Andernfalls wäre die Aussage so zu interpretieren, dass sich Tiger-Objekte aus neun Unterarten zusammensetzen.

Im Hinblick auf die oben diskutierten Befunde wird in der vorliegenden Untersuchung die generische Referenz mit dem definiten Plural als kollektive Gattungsreferenz betrachtet. Die Definitheit des gattungsreferierenden Plurals lässt sich analog zu seinem partikulären Gegenstück begründen, d. h. der gattungsreferierende Plural ist definit, weil er für ein Kollektiv steht, das aufgrund seiner Maximalität einzigartig ist. Der Unterschied zwischen den generischen und partikulären Referenten des definiten Plurals liegt lediglich in der pragmatischen Abgrenzung der letzteren.

Mit Hinblick auf die festgelegten generischen Referenten der definiten NPn im Singular und Plural lassen sich die im Kapitel 2 aufgegriffenen distributiven Unterschiede zwischen den beiden NP-Formen erläutern. Der Übersichtlichkeit halber werden die entsprechenden Beispielpaare im Folgenden repliziert. Die Aussagen (52)–(53) kontrastieren hinsichtlich kollektiver Interpretation:

- (52) *?Der deutsche Käufer* hat letztes Jahr 11000 BMWs gekauft.  
 (53) *Die deutschen Käufer* haben letztes Jahr 11000 BMWs gekauft.

Die Sätze beinhalten ein Prädikat, das kollektiv, d. h. indirekt, auf jeden deutschen Käufer zutreffen muss. Die kollektive Interpretation in (52) scheint jedoch verloren zu gehen. Da der definite Singular auf die abstrakte Gattung referiert, wird der Satz als Aussage über ein abstraktes Individuum, das letztes Jahr 11000 BMWs gekauft hat, verstanden. Dies ist eine nachvollziehbare, aber unübliche Proposition, in der die ursprünglich kollektive Eigenschaft einem abstrakten Individuum zugeschrieben wird. Im Unterschied zum definiten Singular klingt der definite Plural in (53) durchaus natürlich, denn in diesem Fall wird eine kollektive Eigenschaft einem Kollektiv zugeordnet.

---

14 Beispiel in Anlehnung an von Frieling (2002, S. 45).

Aufgrund referentieller Unterschiede zwischen den singularischen und pluralischen Definiten bekommen Generalisierungen über menschliche Gruppen auch verschiedene konnotative Schattierungen. Vgl.:

(54) ?*Der Amerikaner* sieht gerne fern.

(55) *Die Amerikaner* sehen gerne fern.

Wie Radden (2009) anmerkt, ist das Verhalten der Menschen unabhängig von ihrer Zugehörigkeit zu dieser oder jener Gruppe und aufgrund ihrer intellektuellen und geistigen Entwicklung sehr variabel. Der definite Singular in (54) hingegen stellt die Menschen mit amerikanischer Nationalität undifferenziert als abstraktes Individuum dar, etwa wie es bei Tiergattungen der Fall ist, und klingt daher unangemessen (vgl. von Frieling, 2002). Als angemessene Form für Generalisierungen über menschliche Gruppen gilt der definite Plural, der z. B. eine Nationalität nicht als taxonomische Einheit einstuft, sondern als Kollektiv darstellt. Darüber hinaus lässt sich dank der kollektiven Lesart das Implizieren von jeglichen Stereotypen vermeiden. Anders gesagt, bedeutet die Aussage (55) nicht, dass jeder Amerikaner gerne fernsieht, sondern dass es lediglich eine charakteristische Eigenschaft der amerikanischen Nation ist.

Das nächste Beispielpaar repliziert den Fall der *representative object*-Interpretation, bei welcher der definite Plural unter Fragezeichen steht:

(56) Hier wurde in diesem Sommer *der schwarze Schwan* gesichtet.

(57) ?Hier wurden in diesem Sommer *die schwarzen Schwäne* gesichtet.

Die Verwendung des gattungsreferierenden definiten Singulars in (56) klingt angemessen – unabhängig von der Anzahl der Gattungsrepräsentanten, von welchen das Prädikat auf die Gattung projiziert wird. Selbst wenn nur *ein* schwarzer Schwan gesichtet wurde, wurde automatisch die Gattung „Schwarzer Schwan“ gesichtet. Dies scheint, wie (57) illustriert, eine unzureichende Bedingung für die Projektion des Prädikats auf das entsprechende generische Kollektiv zu sein. Anders gesagt, wenn schwarze Schwäne an einem bestimmten Ort gesichtet wurden, bedeutet es keineswegs, dass dies auf alle schwarzen Schwäne kollektiv zutrifft. Solch eine Projektion ist nur in Kontexten mit der *avant-garde*-Interpretation möglich, und zwar wenn ein Objekt bzw. ein Set von Objekten eine für die ganze Gattung höchst relevante Eigenschaft aufweist (siehe Kapitel 2).

## 4 Semantik des indefiniten Singulars und seine Interpretation in generischen Kontexten

Die Interpretationsgrundlage für die Bedeutung des indefiniten Artikels ergibt sich aus den Kontexten, in denen die Verwendung des definiten Artikels ungrammatisch ist. Da der definite Artikel nicht verwendet werden kann, wenn der Referent die Einzigkeitsbedingung nicht erfüllt, scheint die Semantik des indefiniten Artikels durch das Nicht-Einzigkeitsmerkmal geprägt zu sein. Hawkins (1991) weist darauf hin, dass dies eine pragmatische Interpretation des indefiniten Artikels sei; semantisch gesehen sei er neutral. So wird der indefinite Artikel verwendet, wenn in einem P-Set mehr als ein potenzieller Referent vorhanden ist, wie in *Ich habe mir einen Finger gebrochen*, aber auch wenn es im P-Set keinen potenziellen Referenten gibt:

With regard to uniqueness, if there is no mutually manifest P-set for an indefinite description, then there will, in principle, be neutrality to uniqueness (cf. I just saw a unicorn), although knowledge of the world will generally enrich these interpretations in the direction of non-uniqueness (cf. / met a professor once upon a time). (Hawkins, 1991, S. 430)

Die semantische Neutralität des indefiniten Artikels hinsichtlich Einzigkeit erscheint plausibel, jedoch weist sie nicht darauf hin, dass der indefinite Artikel ein semantisch leeres Gegenstück des definiten Artikels darstellt. Während eine definite NP eine pragmatisch oder logisch bedingte Einzigkeit ihres Referenten voraussetzt, wird die indefinite NP systematisch dann und nur dann gebraucht, wenn ihr Referent einer aus der unbestimmten Menge von potenziellen Referenten ist. Wenn auch der indefinite Referent als einer aus der durch das P-Set definierten Menge verstanden wird, ist das eine Interpretation, die, wie Hawkins anmerkt, ausschließlich pragmatisch ist. Diese Beobachtung führt zur Annahme, dass die Semantik des indefiniten Artikels durch die Bedingung der unbeschränkten Menge von potenziellen Referenten geprägt ist. Aus der Perspektive dieser Auffassung lässt sich erklären, warum die Referentenwiederaufnahme mit einer indefiniten NP unmöglich ist. Da eine indefinite NP für den Referenten steht, der als einer aus der unbestimmten Menge von potenziellen Referenten repräsentiert wird, müsste damit für die Wiederaufnahme die entsprechende Menge im vorherigen Diskurs existieren – eine Bedingung, die unter keinen Umständen erfüllbar ist. Die Menge potenzieller Referenten wird immer durch die Anzahl von bereits eingeführten Referenten determiniert, so wie z. B. in (58) durch zwei Referenten:

(58) Ich habe *einen Hund*<sub>1</sub> und *eine Katze*<sub>2</sub>. Ich habe *ein Tier*<sub>3</sub> auf der Straße gefunden.

Aus diesem Grund wird jede Verwendung einer indefiniten NP ausschließlich auf die Einführung eines Referenten hinweisen<sup>15</sup>.

Mit der aufgestellten Annahme zur semantischen Bedingung des indefiniten Artikels sind anschließend die Besonderheiten in der Verwendung des indefiniten Singulars in generischen Kontexten zu betrachten. Im Kapitel 2 wurde darauf hingewiesen, dass der indefinite Singular mit den Gattungsprädikaten unangemessen klingt, weil er keine Gattung kennzeichnen kann. Die vorliegende Artikeltheorie erklärt, warum dies unmöglich ist. Wie festgestellt, steht eine indefinite NP für einen von unbestimmt vielen potenziellen Referenten. Da aber eine Gattung logisch bedingt ein einzigmöglicher Repräsentant des Gattungskonzeptes ist, kann sie nicht mit einer indefiniten NP zum Ausdruck gebracht werden. Demgegenüber ist der Gebrauch des indefiniten Singulars zur Bezeichnung von Untergattungen semantisch angemessen. Das Konzept der Untergattung impliziert mehrere potenzielle Repräsentanten und erfordert somit zur Referenz auf einen davon eine indefinite NP. In diesem Fall ist der indefinite Singular taxonomisch zu lesen (vgl. Dayal, 2004; Krifka et al., 1995). Wenn also die NP *ein Löwe* für eine seiner Untergattungen steht, beispielsweise den Senegal-Löwen, ist die Verwendung des indefiniten Singulars semantisch angemessen:

(59) *Ein Löwe* ist in Westafrika verbreitet.

Wird mit der NP *ein Löwe* auf keine Untergattung referiert, muss sie einen partikulären Referenten haben, d. h. ein Objekt „Löwe“. Aufgrund dieser Interpretation des indefiniten Singulars entsteht jedoch die unrealistische Behauptung, dass ein konkretes Objekt verbreitet ist.

Zu identifizieren ist nun der Grund, aus welchem Charakterisierungsaussagen mit dem objektreferierenden indefiniten Singular als Generalisierungen über Gattungen verstanden werden können. Vgl.:

(60) *Die Eule* jagt in der Nacht.

(61) *Die Eulen* jagen in der Nacht.

(62) *Eine Eule* jagt in der Nacht.

(63) *Eulen* jagen in der Nacht.

Alle vier Aussagen in der Beispielsequenz können als Generalisierungen über die Gattung „Eule“ verstanden werden. Die Sätze (60) und (61) erlauben dies aufgrund des gattungsreferierenden Potentials der definiten NP.

---

15 Diese Überlegungen stimmen mit der *Novelty*-Bedingung von Heim (1982) überein, der zufolge eine indefinite NP immer neu ist, d. h. keine entsprechende Karteikarte im File hat.

Die NP in (62) hingegen ist objektreferierend. Man kann annehmen, dass es das Charakterisierungsprädikat ist, das der rezipierten Generalisierung über die Eulengattung zugrunde liegt. Da Charakterisierungsprädikate auch Eigenschaften von spezifischen Individuen (z. B. Maria, unser Hund, dieser Mann etc.) ausdrücken können, sind sie keine ausreichende Bedingung für Generalisierungen über Gattungen. Es ist anzunehmen, dass die entscheidende Rolle dabei der unspezifischen Lesart der indefiniten NP zukommt, die das Charakterisierungsprädikat neben der spezifischen Lesart erlaubt: Wenn eine arbiträre Eule gefährlich ist, dann muss diese Eigenschaft auf die ganze Eulengattung zutreffen (vgl. Bisle-Müller, 1991; Hawkins, 1978; Radden, 2009). In diesem Fall kann man also auf eine Art indirekte Gattungsreferenz schließen. Es ist zu betonen, dass solch eine Generalisierung bestimmte Restriktionen voraussetzt: Wenn ein einzelnes arbiträres Exemplar die ganze Gattung repräsentiert, müssen alle Exemplare der Gattung bezüglich der im Prädikat ausgedrückten Eigenschaft äquivalent sein. Das bedeutet, dass der unspezifische Referent des indefiniten Singulars nur mit jenen Charakterisierungsprädikaten kombinierbar ist, die eine charakteristische Eigenschaft einer Gattung ausdrücken – eine Eigenschaft, die auf alle typischen Exemplare zutrifft. Welches Prädikat als charakterisierend gelten kann, hängt von der zu beschreibenden Gattung ab. Vgl.:

- (64) Ein Vogel hat einen Schnabel und Federn.<sup>16</sup>
- (65) \*Ein Vogel singt.
- (66) Eine Nachtigall singt.

Die Generalisierung in (64) ist akzeptabel, weil der Prototyp<sup>17</sup> eines Vogels über einen Schnabel und Federn verfügt. Im Gegensatz dazu klingt die Generalisierung in (65) etwas merkwürdig, weil sich viele Vogel-exemplare finden, die Laute von sich geben, die mit dem Vogelgesang nichts zu tun haben, z. B. Uhus, Enten, Hühner etc. (vgl. Radden, 2009). Wenn man sich dagegen auf eine typische Nachtigall bezieht, die für ihren herrlichen Gesang berühmt ist, scheint die Generalisierung über die Gattung „Nachtigall“ mit dem Prädikat *singt* in (66) angemessen zu sein.

---

16 Beispiele (64) und (65) von Radden (2009, S. 288), Übersetzung von mir.

17 Der Prototyp wird hier „considered as a more abstract, generic concept, that was constituted from the different ways in which the category members resembled each other, and differed from non-members“ (Hampton 2006, S. 80).

## 5 Zur Semantik des bloßen Plurals

Zur Semantik der bloßen Plurale gibt es unterschiedliche Meinungen. Laut Carlson (1977) sind bloße NPn immer gattungsreferierend und lassen sich, wie die folgenden Beispiele illustrieren, als „this kind of thing“ (dt. „diese Art Sache“) paraphrasieren. Die nicht-generische Interpretation, die in (68) entsteht, sei durch die episodische Lesart des Prädikates bedingt.

- (67) *Kühe* sind reine Pflanzenfresser. = *Diese Art Tier* ist ein reiner Pflanzenfresser.  
 (68) Auf der Weide liegen *Kühe*. = Auf der Weide liegt *diese Art Tier*.

Nach Krifka (2003) ist die NP in (67) gattungsreferierend, aber die in (68) indefinit, d. h. objektreferierend. Die beiden Referenztypen seien nur Interpretationen von bloßen Pluralen, die an sich für Eigenschaften stehen. Davidse (2009) dahingegen betrachtet gattungsreferierende und indefinite Plurale als zwei unterschiedliche Typen des bloßen Plurals, die grammatische, textuelle sowie konzeptionelle Unterschiede aufweisen. Die Diskussion zur Semantik von bloßen NPn ist sehr umfangreich und erfordert eine tiefgreifende Analyse, die den Rahmen des vorliegenden Artikels sprengen würde. Aus diesem Grund werden die folgenden Ausführungen auf die Skizzierung des Ansatzes von Carlson (1977) sowie theoretische Überlegungen zur Semantik von bloßen Pluralen unter der Indefinitheitsannahme eingeschränkt.

### 5.1 Bloßer Plural als gattungsreferierend

Einen der zentralen Befunde in der Arbeit von Carlson (1977) stellen sprachliche Evidenzen zur Nichtambiguität der bloßen Plurale dar. Carlson beobachtet, dass die NP-Form bei Gattungsprädikaten sowie Charakterisierungsprädikaten systematisch die generische Interpretation erhält, während bei episodischen Prädikaten die partikuläre Interpretation evoziert wird. Dieses Muster lasse darauf schließen, dass die Unterschiede in der Interpretation der bloßen Plurale nicht mit der NP-Semantik, sondern mit ihrem kontextuellen Umfeld zusammenhängen. Als Gegenargument könne man annehmen, dass es die NP-Semantik sei, die den gattungsreferierenden bzw. objektreferierenden NPn nicht erlaubt, in denselben Kontexten aufzutreten, jedoch lasse sich diese Annahme mit den Beispielen folgender Art widerlegen:

- (69) Anna hat *Avocados* gekauft, obwohl *sie* sehr teuer sind.  
 (70) *Avocados* sind sehr teuer, aber Anna hat *sie* trotzdem gekauft.



Nach Carlson habe die NP *Avocados* in (69) eine partikuläre und in (70) eine generische Deutung. Gäbe es die angeblichen semantischen Einschränkungen seitens der NP, so könnte ihr Referent, was auch immer das ist, in dem nachfolgenden generischen bzw. partikulären Kontext nicht wiederaufgenommen werden. Das ist aber, so Carlson, nicht der Fall. Stattdessen zeugen die anaphorischen Beziehungen zwischen dem bloßen Plural und der Proform davon, dass sie für denselben Referenten stehen, was seinerseits demonstriert, dass der bloße Plural keine semantische Ambiguität aufweist.

Was prägt dann die Semantik der bloßen Plurale: Indefinitheit oder Gattungsreferenz, die, wie festgestellt, per se definit ist? Anhand von mehreren Kontexten versucht Carlson zu beweisen, dass der bloße Plural kein pluralisches Gegenstück des indefiniten Singulars sei. Er stellt z. B. fest, dass der bloße Plural bei Universalquantoren oder in opaken Kontexten im Unterschied zum indefiniten Singular nur eine opake Lesart bekommt. Dahingegen habe die NP, mit der die Referenz auf die Gattung overt markiert wird, etwa wie *diese Art Sache*, in den erwähnten Kontexten ein mit dem bloßen Plural analogisches Interpretationsmuster. Zudem illustriert Carlson, dass der bloße Plural mit der overt generischen NP in generischen Kontexten austauschbar sei:

- (71) *Löwen sind Säugetiere.* = *Diese Art Tier ist ein Säugetier.*
- (72) *Löwen haben eine Mähne.* = *Diese Art Tier hat eine Mähne.*
- (73) *Löwen sind in Afrika verbreitet.* = *Diese Art Tier ist in Afrika verbreitet.*

Die 1:1-Überlappung in der Interpretation der beiden NPn führt Carlson zu der Annahme, dass der bloße Plural an sich gattungsreferierend sei. Die Gattungen seien als abstrakte Individuen zu betrachten, und der bloße Plural fungiere als Eigenname dafür. Unter dieser Prämisse eröffnet sich eine interessante Perspektive auf die Interpretation von Aussagen sowohl mit Gattungen als auch mit konkreten Individuen. Als abstraktes Individuum bestehe eine Gattung, so Carlson, aus konkreten Individuen, während ein konkretes Individuum als Gebilde seiner Zustände zu verstehen sei. Die Bestandteile beider Individuentypen nennt Carlson „Stadien“, und die Eigenschaften, die ihnen durch das Prädikat zugeschrieben werden, „Zustände“. Die Hinweise darauf, ob ein Individuum oder sein Stadium prädiziert wird, lassen sich aus der kontextuellen Umgebung, in der sich die NP befindet, ableiten; die NP selbst referiert immer auf ein Individuum. Vgl.:

- (74) *Jack runs vs. is running.*
- (75) *Dogs bark vs. are barking.*

In (74) wird mit dem Prädikat im Present Simple eine charakteristische Eigenschaft von einem Individuum Jack ausgedrückt, während mit dem Present Continuous eine Eigenschaft dem Stadium des Individuums Jack zugewiesen wird. Analoges gilt für die Interpretation des Satzes mit dem bloßen Plural. Im ersten Satz charakterisiert das Prädikat die Hundegattung, im zweiten Satz hingegen nur das Stadium der Gattung, das als ihr Subset verstanden werden kann. Der Logik dieses Ansatzes folgend gibt es im Unterschied zu (76) keine anaphorischen Beziehungen zwischen den NPn in (77), weil der bloße Plural und der definite Plural unterschiedliche Referenten haben: Gattung und konkrete Individuen.

- (76) Auf der Weide liegt *eine Kuh*<sub>1</sub>. *Die Kuh*<sub>1</sub> ist braun.  
 (77) Auf der Weide liegen *Kühe*<sub>1</sub>. *Die Kühe*<sub>2</sub> sind braun.

Dieser Kontrast lässt die Frage zur Festlegung der definiten Referenten aufkommen. Der definite Referent in (76) kann aufgrund seiner expliziten Vorerwähnung mit der indefiniten NP identifiziert werden, sodass mit der NP *die Kuh* auf die Kuh referiert wird, die auf der Weide liegt. In (77) bietet sich eine ähnliche Interpretation des definiten Referenten an, jedoch gibt es, wenn man sich an die Theorie von Carlson hält, keine explizite Einführung seines Antezedens. Man könnte eine implizite Referenteneinführung aufgrund der Erwähnung der Gattung annehmen, was jedoch stellt dieser Referent dar? Ist es ein pluralisches Set oder ein singuläres Objekt? Der Kontext kann dazu keine Hinweise liefern, sofern eine Gattung sowohl durch *ein* als auch mehrere Objekte repräsentierbar ist. Dies wiederum würde bedeuten, dass als definitiver Referent – abhängig davon, wie viele Kühe auf der Weide liegen – sowohl ein Individuum als auch ein Set aus mehreren Individuen fungieren kann. Das ist jedoch nicht der Fall, wie (78) zeigt:

- (78) Auf der Weide liegen *Kühe*<sub>1</sub>. \**Die Kuh*<sub>2</sub> ist braun.

Die Unangemessenheit der singularischen Verwendung lässt darauf schließen, dass der bloße Plural den definiten Referenten determiniert, und das ist schwierig anders zu erklären als aufgrund der anaphorischen Beziehung, in der die beiden NPn zueinanderstehen. Sollte es sich um Koreferenz in (78) handeln, sollte der bloße Plural nicht als gattungsreferierend betrachtet werden, sondern als indefinit. Unter Betrachtung der folgenden Kontexte lassen sich weitere Evidenzen für Indefinitheit gewinnen, vgl.:

- (79) Anna hat *Vögel* gefüttert. Peter hat *sie* (= *diese Art Tier*) auch gefüttert.  
 (80) Peter hat *Bäume* gefällt. Max hat *sie* (≠ *diese Art Sache*) auch gefällt.

In (79) ist die Proform mit der overt generischen NP austauschbar und steht somit nicht im Konflikt mit der Annahme von Carlson. Dahingegen scheint die Proform in (80) eine unterschiedliche Semantik im Vergleich zur overt generischen NP zu haben, d. h. sich genau auf die Bäume zu beziehen, die Peter gefällt hat. Vor dem Hintergrund dieser Befunde kann man wiederum versucht sein, auf die semantische Ambiguität des bloßen Plurals zu schließen. Das ist jedoch nicht die Meinung, die hier vertreten wird. Stattdessen wird angenommen, dass der bloße Plural indefinit ist und dass die generische Interpretation der NP-Form das Ergebnis kontextueller Einflüsse darstellt. Diese Annahme wird im Folgenden diskutiert.

## 5.2 Bloßer Plural als indefinit

Was ist der Referent einer indefiniten pluralischen NP? Mit Hinblick auf den ontologischen Aspekt, der mit der Pluralität des Referenten zusammenhängt, kann er kein Kollektiv darstellen. Im Unterschied zum definiten pluralischen Referenten, dessen Grenzen durch die Maximalität seiner Elemente bestimmt sind, kann der indefinite pluralische Referent aufgrund seiner unbestimmten Elementenanzahl nicht abgegrenzt werden und folglich nicht als Einheit repräsentiert werden. Es wird deswegen angenommen, dass der indefinite Referent eine Menge ist. Mit Hinblick auf die Indefinitheit muss die Menge eine von mehreren möglichen Mengen darstellen (siehe Kapitel 4). Diesen theoretischen Überlegungen zufolge stellt der indefinite pluralische Referent eine – unter mehreren möglichen – Menge mit unbestimmter Elementenanzahl dar.

Mit der vorliegenden Auffassung ist zunächst auf die Interpretation des indefiniten Plurals in Aussagen mit Charakterisierungs- und episodischen Prädikaten einzugehen. Aufgrund ihrer Unbestimmtheit kann die Anzahl der Mengenelemente von einem bis alle variieren. Da die Mengengrenzen nicht festgelegt werden können, kann angenommen werden, dass die mit dem Charakterisierungsprädikat ausgedrückte Eigenschaft auf alle möglichen Elemente der Menge zutrifft, und folglich, dass der indefinite Plural als allquantifizierend und in diesem Sinne als generisch zu interpretieren sei. Semantisch gesehen aber kann die Charakterisierungsaussage wie *Löwen haben eine Mähne* auch wahr sein, wenn das Prädikat auf alle, die meisten oder auch wenige Löwen zutrifft. Dasselbe Prinzip liegt den Kontexten zugrunde, in denen der indefinite Plural als partikulär interpretiert wird. Dass die Menge von Schwänen in *Hier wurden in diesem Sommer schwarze Schwäne gesichtet* wahrscheinlich nicht aus allen schwarzen Schwänen der Welt besteht, ist keine semantische, sondern eine pragmatische Folgerung.

Was den indefiniten Plural als Subjekt der Gattungsprädikate betrifft, so bekommt er eine taxonomische Lesart, d. h. der indefinite Plural steht für

die Menge, deren Elemente als die Unterarten der entsprechenden Gattung verstanden werden sollen. Vgl.:

(81) *Die Dinosaurier* sind ausgestorben.

(82) *Dinosaurier* sind ausgestorben.

Mit der Aussage (81) wird behauptet, dass das Kollektiv, dessen Grenzen mit dem letzten Dinosaurier enden, ausgestorben ist. Anders gesagt, es gibt das Kollektiv nicht mehr. Wenn man jedoch einem Dinosaurier begegnet, ist der Satz nicht mehr wahr – das Kollektiv, selbst wenn es nur einem Objekt besteht, gibt es doch. Dem Satz (82) liegen andere Wahrheitsbedingungen zugrunde. Der indefinite Plural scheint sogar die Existenz von ganzen Unter-gattungen zu erlauben (vgl. Laca, 1992; Schaden, 2003). Diese Toleranz gegen Ausnahmen lässt sich mit der hier vertretenen Annahme einordnen, der zufolge die Menge „Dinosaurier“ als Referent des indefiniten Plurals aus einer unbestimmten Anzahl von Dinosaurierunterarten besteht. So muss das der Menge zugeordnete Prädikat nicht auf alle Unterarten zutreffen.

Die Annahme, dass der Referent einer indefiniten pluralischen NP eine Menge mit unbestimmter Elementenanzahl darstellt, erklärt, warum der indefinite Plural bei manchen abgeleiteten Gattungsprädikaten unangemessen ausfällt. Als Menge ist der indefinite pluralische Referent für nicht-realistische Aussagen verantwortlich, die sonst bei definiten Pluralen eine *internal-comparison*- und eine *collective-property*-Interpretation erhalten. Vgl.:

(83) *!Wölfe* werden größer, wenn wir Richtung Norden fahren.

(84) *!Deutsche Käufer* haben letztes Jahr 11000 BMWs gekauft.

Bei der definiten NP-Verwendung bedeutet die Aussage (83), dass das Kollektiv „Wölfe“ mit sich selbst in nördlicheren Teilen verglichen wird. Wenn man versucht, die Aussage mit dem indefiniten Plural zu interpretieren, würde sie bedeuten, dass Wolfexemplare die Eigenschaft haben, größer zu werden, wenn man Richtung Norden fährt. Eine solche Interpretation entspricht dann aber nicht der Funktion der *internal comparison*-Prädikation und stellt ein surrealistisches Bild dar, das am Rande der semantischen Plausibilität steht. Auch die Unglaubwürdigkeit der Aussage (84) kommt davon, dass der indefinite pluralische Referent nicht als Kollektiv, sondern als Menge von einzelnen deutschen Käufern repräsentiert wird, weshalb die Eigenschaft, letztes Jahr 11000 BMWs gekauft zu haben, sich als distributiv erweist.

Soweit scheinen partikuläre und generische Kontexte eine plausible Erklärung unter der Betrachtung des bloßen Plurals als indefiniter NP zu finden. Anschließend wenden wir uns dem Phänomen der anaphorischen Beziehungen zu, das nach Carlson als einer der wichtigsten Belege für die

Gattungsreferenz des bloßen Plurals gelten sollte. Grundsätzlich sind die anaphorischen Beziehungen nur dann plausibel, wenn das Antezedens und seine Proform für ein und denselben Referenten stehen. Im Falle des indefiniten Plurals geht es um Koreferenz dann, wenn als Referent dieselbe Menge fungiert. Zu betrachten ist nun das folgende Beispiel:

- (85) Anna hat *Vögel* gefüttert. Peter hat *sie* (*die Vögel* vs. *Vögel*) auch gefüttert.

In der Aussage (85) hat das Pronomen *sie* zwei Interpretationen. Entweder steht das Pronomen genau für die *Vögel*, die Anna gefüttert hat, und ist mit der definiten NP *die Vögel* austauschbar, oder das Pronomen ist informativ schwächer, d. h. gibt keine Hinweise darauf, welche *Vögel* Peter gefüttert hat, und kann durch die indefinite NP *Vögel* ersetzt werden. Beide Interpretationen weisen Parallelen bei entsprechenden singularischen Ausdrücken auf:

- (86) Anna hat *einen Vogel* gefüttert. Peter hat *ihn* vs. *einen* auch gefüttert.

Im Unterschied zum indefiniten Plural besitzt der indefinite Singular für jede der Interpretationen unterschiedliche pronominale Formen. Das Personalpronomen wird für die Referenz auf den bereits bekannten Vogel verwendet, während das indefinite Pronomen für einen nicht näher definierten Vogel steht. In dieser Hinsicht kann das indefinite Pronomen als singuläres Gegenstück des Pronomens *sie* in der indefiniten Interpretation verstanden werden. Das folgende Beispielpaar zeigt, dass der Gebrauch von pluralischen indefiniten Pronomen gegenüber den singularischen und den definiten eingeschränkter ist:

- (87) Peter hat *einen Baum* gefällt. Max hat *ihn* (*den Baum*) vs. *einen* (*einen Baum*) auch gefällt.  
 (88) Peter hat *Bäume* gefällt. Max hat *sie* (*die Bäume* vs. *?Bäume*) auch gefällt.

Ausgehend von den vorliegenden Beispielen kann man annehmen, dass der Referent eines indefiniten Pronomens den Default-Eigenschaften seines nominalen Gegenstücks entsprechen muss. Beim indefiniten Singular ist die Existenz solch eines Referenten immer möglich, d. h. wenn Peter einen Baum gefällt hat, ist die Existenz eines weiteren Einzelobjektes, das ein Baum ist und von Max gefällt werden kann, nicht ausgeschlossen. Beim indefiniten Plural ist es anders. Die oberste Grenze der Menge von ‚fällbaren‘ Bäumen liegt bei Max wegen Peters Fällens tiefer, als es bei Peter der Fall war. So

kann Max' Menge von gefälltten Bäumen nicht den Default-Eigenschaften des indefiniten pluralischen Referenten entsprechen, d. h. die Variation der Elementenanzahl von einem bis alle ist ausgeschlossen, und kann mit dem indefiniten Pronomen nicht kennzeichnet werden. Der Grund, aus welchem bei der gleichen syntaktischen Struktur wie in (85) die Referenz mit *sie* sich nur als definit erweist, hängt mit der Verbsemantik zusammen. Das Verb *fällen* gehört zu den faktiven Verben, d. h. denjenigen Verben, die das Versetzen eines Objektes in einen neuen Zustand bezeichnen. Das bedeutet, dass die Bäume, die von Peter gefällt wurden, nicht mehr von Max gefällt werden können. Dahingegen hat die Semantik des Verbes *füttern* keinen Einfluss auf die potenzielle Elementenanzahl der Menge von Vögeln, die gefüttert werden können. Wenn ein Vogel gefüttert wird, kann er nochmal gefüttert werden. Damit werden die Default-Eigenschaften des indefiniten pluralischen Referenten aufrechterhalten und die Möglichkeit zur Referenz darauf mit dem indefiniten Pronomen gegeben. Mit Hinblick auf diese Beobachtungen ist im aktuellen Zusammenhang die pronominale Verwendung in Sätzen wie (89) nicht als Indiz für die generische Koreferenz zu betrachten, sondern als Ausdruck der indefiniten pluralischen Referenz, die nicht im Konflikt mit der Verbsemantik oder sonstigen möglichen semantischen Merkmalen der Satzkonstituenten steht:

(89) *Avocados* sind sehr teuer, aber Anna hat *sie* (*Avocados*) trotzdem gekauft.

Die Tatsache, dass einer Menge von Avocados die Eigenschaft, teuer zu sein, zugewiesen wird, hat keinen semantischen Einfluss auf die Menge von Avocados mit der Eigenschaft, von Anna gekauft worden zu sein. d. h., die Anzahl der von Anna gekauften Avocados kann von einem bis alle variieren. Diese Menge stimmt somit mit den Default-Eigenschaften des indefiniten pluralischen Referenten überein und lässt sich mit dem indefiniten pluralischen Pronomen bezeichnen.

## 6 Zusammenfassung

In diesem Beitrag wurde die Frage nach den Gründen für NP-Restriktionen in generischen Kontexten vor dem Hintergrund der Artikelsemantik aufgegriffen. Ausgehend von der Einzigkeitstheorie zur Bedeutung des definiten Artikels konnte gezeigt werden, dass die (In)Definitheit grundlegende Eigenschaften von generischen Referenten prägt. Die definite NP steht im Singular für eine abstrakte Gattung und im Plural für ein generisches Kollektiv, weil die beiden Referententypen logisch bedingt den einzig möglichen

Repräsentanten des nominalen Konzeptes darstellen. Aus demselben Grund kann die indefinite NP, die einen von unbestimmt vielen potenziellen Referenten kennzeichnet, nicht gattungsreferierend sein. Die generische Interpretation von Charakterisierungsaussagen mit indefiniten singularischen NPn kann nur aufgrund ihrer Unspezifität erreicht werden. In Kontexten mit Prädikantentypen, die eine spezifische Lesart der NPn voraussetzen, lässt sich keine generische Interpretation ableiten. Einzuschränken ist, dass der indefinite Singular eine taxonomische Interpretation erhalten kann, in der er eine der Untergattungen einer Gattung bezeichnet. Die Semantik der bloßen Plurale konnte hier unter der Indefinitheitsprämisse beschrieben und begründet werden. Es wurde festgestellt, dass indefinite Plurale für Referenten stehen, die sich als Mengen mit unbestimmter Elementenanzahl manifestieren. Aufgrund dieser Unbestimmtheit können Aussagen mit Charakterisierungsprädikaten als Aussagen über alle möglichen Elemente der entsprechenden Menge rezipiert werden und folglich zu einer generischen Interpretation führen. Mit Gattungsprädikaten erwerben indefinite Plurale, ähnlich dem indefiniten Singular, eine taxonomische Lesart, in der die Menge aus einzelnen Untergattungen besteht. Die Semantik der indefiniten Plurale konnte hier nicht umfangreich behandelt werden, deswegen ist ein systematischer Vergleich der gewonnenen Einblicke mit den Erkenntnissen der einschlägigen Forschung wünschenswert.

## Danksagung

Dieser Beitrag entstand als überarbeitete Fassung des ersten Kapitels meiner Masterarbeit „Der deutsche Artikel als Referenzmarker in generischen Kontexten. Eine empirische Studie zur rezeptiven Kompetenz bei ukrainischen Lernern des Deutschen als Fremdsprache“. Für detaillierte und hilfreiche Kommentare zu der Vorversion dieses Beitrags möchte ich mich herzlich bei Michael Herweg bedanken.

## Literaturverzeichnis

- Barton, D. (2016). *Generische Nominalphrasen bei deutsch-französischer Zweisprachigkeit: Zur Verwendung des Definitartikels bei erwachsenen Herkunftssprechern* (Dissertation). Universität Hamburg. [urn:nbn:de:gbv:18-80359]
- Behrens, L. (2005): Genericity from a Cross-Linguistic Perspective. *Linguistics* 43(2), 275–344.
- Bisle-Müller, H. (1991). *Artikelwörter im Deutschen*. Niemeyer.
- Blühdorn, H. (2001). Generische Referenz. Ein semantisches oder ein pragmatisches Phänomen? *Deutsche Sprache* 29, 1–19.

- Carlson, G. (1977). A unified analysis of the English bare plural. *Linguistics and Philosophy* 1(3), 413–457.
- Christophersen, P. (1939). *The Articles: A Study of Their Theory and Use in English*. Einar Munksgaard.
- Chur, J. (1993). *Generische Nominalphrasen im Deutschen. Eine Untersuchung zu Referenz und Semantik*. Niemeyer.
- Davidse, K. (2009). *Bare plural NPs in English: indefinites versus generics*. [[https://www.researchgate.net/publication/305508293\\_Bare\\_plural\\_NPs\\_in\\_English\\_indefinites\\_versus\\_generics](https://www.researchgate.net/publication/305508293_Bare_plural_NPs_in_English_indefinites_versus_generics)]
- Dayal, V. (2004). Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy* 27(4), 393–450.
- von Frieling, S. (2002). *Generische Syntagmen im Französischen und Deutschen. Ein Beitrag zur menschlichen und maschinellen Übersetzung*. Lang.
- Gerstner-Link, C. (1995). *Über Generizität: Generische Nominalausdrücke in singulären und generellen Aussagen*. Fink.
- Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Hrsg.), *Syntax and Semantics 3: Speech Acts* (S. 41–58). Academic Press.
- Grimm, J. (1986). *Untersuchungen zum Artikelgebrauch im Deutschen*. Verlag Enzyklopädie.
- Hampton, J. A. (2006). Concepts as prototypes. *Psychology of learning and motivation* 46, 79–113.
- Hawkins, J. A. (1978). *Definiteness and Indefiniteness: A Study in Reference and Grammaticality Prediction*. Croom Helm.
- Hawkins, J. A. (1991). On (in)definite articles: implicatures and (un)grammaticality prediction. *Journal of Linguistics* 27(2), 405–442.
- Heim I. R. (1982). *The Semantics of Definite and Indefinite Noun Phrases in English* (Dissertation). University of Massachusetts at Amherst. [<https://scholarworks.umass.edu/dissertations/AAI8229562>]
- Ionin, T., Montrul, S., Kim, J.-H., Philippov, V. (2011). Genericity distinctions and the interpretation of determiners in second language acquisition. *Language Acquisition* 18(4), 242–280.
- Kefer, M. (2011). Kollektive und distributive Generizität. In J. Strässler (Hrsg.), *Tendenzen europäischer Linguistik: Akten des 31. Linguistischen Kolloquiums, Bern 1996* (S. 109–113). De Gruyter Mouton.
- Krifka, M. (2003). Bare NPs: Kind-referring, Indefinites, Both, or Neither? *Semantics and Linguistic Theory* 13, 180–203.
- Krifka, M., Pelletier, F. G., Carlson, G. N., ter Meulen, A., Link, G., Chierchia, G. (1995). Genericity. An Introduction. In G. N. Carlson & F. G. Pelletier (Hrsg.), *The generic book* (S. 1–124). University of Chicago Press.
- Levinson, S. C. (1987). *Pragmatics*. Cambridge University Press.
- Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic* 8(1), 339–359.



- Lyons, C. (1999). *Definiteness*. Cambridge University Press.
- Molnár, K. (2014). *Substantivdetermination im Deutschen und im Ungarischen*. De Gruyter.
- Mumm, P.-A. (1995). Generische Bezeichnung. Onomasiologische Aufgaben und ihre Lösungen durch das neuhochdeutsche Artikelsystem. *Sprachwissenschaft* 20(4), 420–467.
- Radden, G. (2009). Generic reference in English: A metonymic and conceptual blending analysis. In K. Panther, L. Thornburg & A. Barcelona (Hrsg.), *Metonymy and metaphor in grammar* (S. 199–228). John Benjamins.
- Russell, B. (1905). On denoting. *Mind* 14(56), 479–493.
- Schaden, G. (2013). Two ways of referring to generalities in German. In A. Mari, C. Beyssade & F. de Prete (Hrsg.), *Genericity* (S. 157–175). Oxford University Press.
- Sperber, D. & Wilson, D. (1986). *Relevance: communication and cognition*. Basil Blackwell.
- Vater, H. (1997). Determination im Deutschen und Polnischen. In G. Koniuszaniec & I. Prokop (Hrsg.), *Festschrift für Andrzej Z. Bzdęga zum 70. Geburtstag* (S. 215–229). Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza, Poznań.



---

Michael Herweg

# Varieties of Manner of Motion: A Frame-Semantic Account

---

**Abstract** This article introduces a fine-grained distinction of types of manners of motion in a frame-semantic setting. The frame-theoretical backdrop is provided by the FAMEu approach presented in Herweg (2020) and augmented with a theory of verbal types and dynamic interpretation in Herweg (2021). FAMEu features an extension of the Cognitive Grammar notion of profiling (Langacker, 2013). The enhanced profiling mechanism allows each type of motion verb to impose its characteristic profile on a basic *LOCOMOTION* frame, which provides an expressive structure that facilitates the representation of major elements of the linguistic description of motion events. The particular profile associated with a motion verb may select, deselect and augment substructures of this general frame, as well as elaborate on significant relations between different substructures. Frame representations for three forms of complex manner constructions and six basic types of manner-of-motion verbs are defined on this theoretical basis and illustrated with examples from German, French and English.

**Keywords** motion verbs; manners of motion; frame semantics; profiling

## 1 The FAMEu approach to the semantics of motion expressions

This article deals with some fine-grained distinctions we observe in the area of manner-of-motion verbs (MoMVs) and constructions. It proposes a system of types of manner of motion that is intended to cover some important characteristics of major classes of MoMVs. The propounded analyses are cast in a frame-semantic setting, namely the FAMEu (short for “A Frame-semantic Account of Motion Expressions with Underspecification”) format, which was set forth in Herweg (2020) and which has been enhanced with a theory of verbal types in the motion domain in Herweg (2021). The core element of FAMEu is the *LOCOMOTION* frame, which provides an expressive structure that allows to represent major elements of the linguistic description of motion

events. Rather than reiterating the details of the FAMEu *LOCOMOTION* frame, and to avoid redundancy, I will only briefly sketch the main elements of the approach and refer the reader to the above-mentioned papers, as well as to the explanations of the sample frames in the forthcoming sections of the present paper, for further details.

A crucial feature of the *LOCOMOTION* frame are the elaborate structures which represent the two fundamental event layers that are effective in motion descriptions, namely the path layer and the manner layer (cf. Gerwien & von Stutterheim, 2016; von Stutterheim et al., 2020). The path layer is represented by a *PATH* attribute,<sup>1</sup> which is an internally complex frame structure that covers different conceptual aspects of the path which the moving entity (i.e., the *FIGURE* in Talmy's terms; cf. Talmy, 1983) covers during its motion. These aspects of paths are represented under two main complex attributes: *ROUTE* and *P\_DIM*. *ROUTE* stands for the path as a complex object comprising an initial (*INIT*), an intermediate (*INTER*) and a final (*FIN*) segment. These path segments may, for instance, be further described by directional complements such as 'source', 'route' and 'goal' PPs, which specify a local relation (*L\_REL*) of the moving entity relative to a ground object (*GROUND*; cf. Talmy, 1983). The *P\_DIM* attribute covers several types of dimensional properties of paths, namely their direction in three-dimensional space (attribute *DIR*), the gestalt / form properties of the pure spatial trace of a path (*SHAPE*), their velocity and throughput (*VEL*) and their spatial extent (*EXTENT*).

The *MANNER* attribute describes the manner in which the figure performs the motion. Motion verbs may express manifold manifestations of manners, particularly in languages with a rich inventory of MoMVs. Verbs like *to walk*, *to run*, *to jog*, *to stride* and *to strut* each mark a specific gait that the moving entity performs. By contrast, verbs like *to hurry*, *to rush*, *to hasten* and *to flee*, which also occur in genuine motion contexts (*to hurry/rush/hasten/flee into the great wide open*), are not inherently associated with any specific gait. These actions can in fact be executed in various modes of movement (by foot, bicycle, roller-skates, car, train, ...). Intuitively, these verbs do not designate a motion that can primarily be characterized by specific bodily activities but rather focus on an expeditious change of location, often additionally associated with a presumed state of mind of the actor. In other words: While the first group of verbs designates manners which can to a large extent be characterized in terms of (the progression of) spatial configurations of particular body parts, the manners expressed by the second group of verbs have a predominantly

---

1 The proposed frames are formally represented as recursive typed attribute-value matrices (cf. Petersen, 2015). Attributes in FAMEu are written in upper case and values in lower case letters; types of values and frames, such as the *LOCOMOTION* type and its subtypes, are written in italic small caps.

temporal – and, as I will argue in section 3.3, often an intention-, emotion- or attitude-related – nature.

In the approach put forward in the present study, distinctions among manners are captured by the types of the frames associated with the respective verbs. The type *S\_MOM* applies to verbs such as *to walk*, *to run*, *to limp*, i.e., verbs whose manner can be predominantly associated with the spatial dimension of the motion because they primarily describe specific gaits. If the manner is predominantly temporal in nature, the corresponding frame receives the type *T\_MOM*. This type applies to verbs that do not focus on specific gaits but primarily highlight time-related features of a locomotion, such as a particularly fast (*to race*, *to rush*, *to dash*) or slow (*to ramble*, *to stroll*) movement. A third type of manner that will play a role in the representation of particular verbs is the exertion of a force (type *F\_MOM*) against an overt or tacit object that serves as an antagonist in the activity, like in *to climb* (cf. Geuder & Weisgerber, 2008). Further types of manner that will be addressed in the present study are *I\_MOM* for motion which relies on an instrument, like in the case of *to bike* and *to sail*, *M\_MOM* for motion that is performed in a specific medium of localization, as with *to swim* and *to fly*, and *MC\_MOM* for verbs like *to hasten* and *to hustle*, which allude to a particular mental constitution or state of mind of the moving entity. We leave it open to future research to determine if further semantically significant types of manners have to be assumed in the motion domain.

The FAMEu *LOCOMOTION* type furthermore includes means to represent and calculate the aspectual class, aspect, tense and situational (deictic) properties of motion expressions. The relevant information is collected under a separate *EVENT\_PROPERTIES* attribute, with subattributes *ASPECTUAL CLASS*, *ASPECT*, *TENSE* and *ORIGO*. These attributes play a minor role in the present work, so I will elaborate on them only as needed to follow the examples discussed in subsequent sections.

The key concept of the frame-semantic analyses of motion expressions in typologically divergent languages (German, French, Russian, Korean, Thai) introduced in Herweg (2020) is an extension of the notion of profiling, as originally developed in Cognitive Grammar (see, e.g., Langacker, 1983, 1987, 2013). In the FAMEu approach, each motion verb imposes its characteristic profile on the *LOCOMOTION* frame as part of its lexical semantics. The particular profile associated with a motion verb – which will be marked via boldface font for ease of presentation<sup>2</sup> – may select, deselect and augment substructures of

---

2 I already used the simple boldface notation for profiled parts of frame representations in Herweg (2020). In a more elaborate formal representation, the profiled frame elements would be linked via indices to a dedicated attribute-value substructure of the frame.

the general frame, such as FIGURE, PATH, MANNER and subattributes thereof, as well as elaborate on significant relations between different substructures. This enhanced FAMEu notion of a verb's profile provides a close-up view of the relevant elements of a frame. It thereby allows for representing subtle differentiations within groups of verbs which otherwise look semantically akin from a broader point of view. Simplifying the analysis to coarse high-level distinctions such as – in the most extreme form – a simple dichotomy between path verbs (like *to approach*) and manner verbs (like *to tiptoe*) risks oversimplifying and potentially blurring these finely nuanced distinctions.

Herweg (2021) complements this approach with a fine-grained system of verbal types which are defined in terms of a variant of dynamic logic, namely the Dynamic Interval Temporal Logic (DITL) of Mani & Pustejovsky (2012). The primary focus of the type system, as explored in Herweg (2020) utilizing the profiling mechanism, lies in distinguishing between path-generating and non-path-generating motion verbs. This differentiation only partially intersects with the standard distinction between path and manner verbs. The path-generating class comprises verbs which designate a motion that conceptually incorporates the incremental composition of a structured path, defined by the changing regions occupied by the moving entity. The verbs of this type may express the notion of directedness (e.g., *to approach*, *to come*, *to sink*; type *MOVE\_PD*), or they may be unspecific about any directedness of the path (e.g., MoMVs in languages like English, such as *to walk*, *to run*, *to creep*; type *MOVE\_P*). Pronounced representatives of the class of non-path-generating motion verbs (verbs of the negative type *~MOVE\_PD*) are MoMVs in languages such as Korean (e.g., *tallida* 'run', *kelda* 'walk') and Thai (e.g., *dəən* 'walk', *wīng* 'run'), which express nothing but the pure activity of executing a particular manner of motion and which need to be accompanied by an obligatory path-generating verb, such as a deictic motion verb, in a serial verb construction.

The present study is a sequel to both Herweg (2020) and Herweg (2021) that is mainly concerned with type distinctions among the values in the MANNER attribute of the *LOCOMOTION* frame. It enhances the FAMEu formalism with additional elements that shall serve to capture the characteristics of the subclasses of MoMVs outlined above. However, before we turn to the varieties of MoMVs as such, we need to look at different manifestations of external manner specifications, i.e., manner specifications which are not expressed by the head verbs of the motion constructions themselves but rather by additional adjuncts that modify the respective verbal heads. This preparatory step serves to introduce some pertinent distinctions and formal devices, which will prove beneficial for the subsequent elaborations on the MoMVs themselves.

## 2 Head-verb external specification of manner

### 2.1 Manifestations of head-verb external manner specification

It is a well-known fact that path-dominant languages such as Romance languages typically (but not always, cf. the recap in Hendriks et al., 2021, p. 2) express the manner of a motion not in the main verb but rather in an adjunct to the path verb (or the VP headed by the path verb), which is composed of an infinite verb form like the French *gérondif* and the Spanish and Italian *gerundio*:

- (1) a. [Fr.] Il sortit de la maison en courant.  
           ‘he exited the house walking / running’  
       b. [Sp.] La botella entró a la cueva flotando.  
           ‘the bottle entered the cave floating’  
       c. [It.] La barca passò sotto il ponte galleggiando.  
           ‘the boat passed under the bridge swimming / floating’

Similar manner specifications external to the finite head verb can be found in non-path-dominant languages such as German. The following examples show a variety of mode and manner specifications by means of participles (participle I), adverbs and adverbial phrases, ranging from detailed characteristics of the gait to concurrent autonomous activities:

- (2) Er ging <...> aus dem Haus. ‘He went / walked out of the house ...’  
       a. hinkend / schwankend           ‘limping / wobbling’  
       b. schnell                           ‘quickly’  
       c. leise                             ‘quietly’  
       d. mit festem Schritt           ‘with a firm step’  
       e. auf Zehenspitzen           ‘on tiptoe / tiptoeing’  
       f. in Pantoffeln                 ‘in slippers’  
       g. kopfschüttelnd / pfeifend   ‘shaking his head / whistling’

(2.a) describes specific gaits by means of the participle I form of manner-of-motion verbs that could also occur as the head verbs of the motion description (*Er hinkte / schwankte aus dem Haus*. ‘He limped / wobbled out of the house’). (2.a) closely corresponds to the Romance constructions in (1). (2.b) characterises the speed with which the motion is executed and (2.c) describes a particular mode that serves to avoid noise. (2.d) and (2.e) further narrow down the gait of the figure, whereas (2.f) does so only implicitly by means of referring to the figure’s particular footwear. And finally, the

participles in (2.g) describe additional actions which the figure performs while executing the motion.

The specifications in (2.a)–(2.f) are all directly (as in [2.a]–[2.e]) or indirectly (as in [2.f]) related to the way the motion is performed by the figure. By contrast, (2.g) describes concurrent accompanying activities which are in principle independent from the motion itself. It does so, however, by the same linguistic means as the additional manner descriptions in (2.a). (2.a) and (2.g) will therefore be represented in a structurally similar way, i.e., under the same core attribute in the *LOCOMOTION* frame – the *FIGURE* attribute –, differentiated only by specific subattributes which expand this attribute. In general, the differences between the various forms of mode and manner characterizations observed in (2) will be reflected by the particular types of attributes that are profiled by the expressions under consideration and how these attributes are related to the core attributes of the *LOCOMOTION* frame. The different mode and manner characterizations will either be represented as additional contributions to the *MANNER* and *PATH* attributes or will be treated as independent thereof, more specifically as attributes directly linked to the figure of the motion.

## 2.2 Infinite manner verbs and structure sharing

The participle I verb forms in (2.a) and (2.g) receive structurally similar representations, distinguished by specific subattributes. The attributes of the figure in the *LOCOMOTION* frame will be expanded beyond those needed for the nominal head and its immediate modifiers. The meanings of the head noun and its immediate modifiers (like in *black cat*, *cat on a hot tin roof*) will be presented under an attribute *PROPERTIES* that expands the *FIGURE* attribute. For mode and manner specifications like *hinkend* ‘limping’ and *pfeifend* ‘whistling’ we will utilize an additional *FIGURE*-related attribute *BEHAVIOUR*. We thus assume the basic expansion of the *FIGURE* attribute shown in figure 1, where *PROPERTIES* is meant to record static (persistent or transient) features assigned to the figure, while *BEHAVIOUR* serves to capture the figure’s dynamic manifestations, most notably qualities of or linked to the activities in which the figure is engaged.

Subattributes of the *PROPERTIES* and *BEHAVIOUR* attributes will be introduced in the following sections. Section 2.4 on depictive manner predicates will deal with further expansions of the *PROPERTIES* attribute. The content of verb-external manner-of-motion specifications like *hinkend* ‘limping’ and *schwankend* ‘wobbling’ will be represented under the *BEHAVIOUR* attribute by means of an additional subattribute *MOM\_SPEC*, which is short for “manner-of-motion specification” and whose value will be transferred via structure sharing – as defined for feature-structure formalisms with



<i>LOCOMOTION</i>	
FIGURE	PROPERTIES
	BEHAVIOUR

Figure 1. The basic structure of the FIGURE attribute

<i>GEHEN</i>			
FIGURE	<b>BEHAVIOUR</b>	MOM_SPEC	$\textcircled{1}$ <b>hinken_gait</b>
EVENT_LAYERS	MANNER	gehen_gait $\sqcup$ $\textcircled{1}$	

Figure 2. The relevant frame extract for *hinkend (gehen)* ‘(go / walk) limping’

<i>GEHEN</i>			
FIGURE	<b>BEHAVIOUR</b>	CONC_ACT	<b>pfeifen</b>
EVENT_LAYERS	MANNER	gehen_gait	

Figure 3. The relevant frame extract for *pfeifend (gehen)* ‘(go / walk) whistling’

unification (cf. Carpenter, 1992, p. 37; Müller, 2013, pp. 145 f.) and indicated by coindexing the respective values with ‘ $\textcircled{1}$ ’ – to the main MANNER attribute of the *LOCOMOTION* frame; see figure 2. Verb-external specifications of concurrent but not manner-of-motion related activities like *pfeifend* ‘whistling’ and *kopfschüttelnd* ‘shaking one’s head’ will also be represented under the BEHAVIOUR attribute, though by means of a special subattribute CONC\_ACT (short for “concurrent activity”). Note that there is no transfer of the content of this supplemental attribute to any of the core attributes MANNER or PATH; see figure 3.

Only the profiles of the participle verbs are marked in boldface in these partial frames. The MANNER value ‘gehen\_gait’ in figure 2 represents the gait in which some part of the foot is always in contact with the ground (cf. Diersch, 1972, p. 55).<sup>3</sup> For the purpose of the present study it suffices to assume that ‘gehen\_gait’ represents the minimum set of properties which distinguishes this particular gait from other gaits such as running, which involves a short aerial phase. The merging of ‘gehen\_gait’ with ‘hinken\_gait’ by means of the merge operation  $\sqcup$ , ‘gehen\_gait  $\sqcup$   $\textcircled{1}$ ’ – or rather, after resolving the reference of  $\textcircled{1}$ , ‘gehen\_gait  $\sqcup$  hinken\_gait’ – further constrains this general gait to an

3 There are uses of *gehen* which do not refer to a specific gait, like in *zum Arzt gehen* ‘go to the doctor’, where no particular manner of motion is expressed. In these cases, *gehen* ‘go’ could rather be represented by a general motion predicate type *MOVE*.

<i>SORTIR</i>					
FIGURE	<i>AGENT</i>				
EVENT_LAYERS	<b>PATH</b>	<b>ROUTE</b>	<b>INIT</b>	<b>L_REL</b>	<b>in</b>
				GROUND	house
			<b>INTER</b>	<b>L_REL</b>	<b>~in</b>
				GROUND	house
EVENT_PROPERTIES	ASPECTUAL_CLASS	...			
	ORIGO	...			
	TENSE	...			
	ASPECT	...			

Figure 4. The representation of French *sortir de la maison* in the (simplified) *LOCOMOTION* frame

impaired and unbalanced manner of motion. The representation of this specific gait will be revisited in the discussion of verbs like *to limp* in section 3.1.

The fact that the activities represented as *MOM\_SPEC* and *CONC\_ACT* are concurrent with the main motion activity, i.e., the activity that determines the type of the frame itself, does not have to be represented separately by specific attribute structures. The concurrence rather follows from the fact that there is one unique temporal anchoring of the entire frame, which is represented under the *EVENT\_PROPERTIES* attribute structure. In conjunction with an appropriate interpretation of the *MOM\_SPEC* and *CONC\_ACT* attributes, the *EVENT\_PROPERTIES* structure properly determines the temporal interpretation of the entire frame and all of its components (cf. Herweg, 2020, 2021).

The approach in figure 2 can be carried over to verb-external manner-of-motion specifications by means of the *gérondif* or *gerundio* in Romance languages like in (1). Assuming that the frame representation of French *sortir* profiles the initial and intermediate segments of the path and declares for them a simple change-of-state in terms of two contrary predications “being inside / at the ground” vs. “not being inside / at the ground” (cf. Herweg, 2020), the relevant extract of the frame representation of *sortir de la maison* looks as shown in figure 4.<sup>4</sup>

Adding a *gérondif* manner-of-motion specification like *en courant* yields the (partial) frame representation in figure 5, which is similar to the one in figure 2. The frame reflects the assumption that the *gérondif* activates the

4 ‘~’ indicates negation. The referential mode of the NP *la maison* is not captured in the frame in figure 4, since the frame-representation of (in)definiteness and quantification is not in the scope of this study. Note that all sample frames in the present article will only show the parts relevant to the respective examples.

<i>sortir</i>					
FIGURE	BEHAVIOUR		MOM_SPEC		Ø <i>courir_gait</i>
EVENT_LAYERS	MANNER	①			
	PATH	ROUTE	INIT	L_REL	in
				GROUND	house
			INTER	L_REL	~in
				GROUND	house

Figure 5. The representation of French *sortir de la maison en courant* (the profile of the gérondif is visualized by a dotted box)

manner layer of the locomotion event concept by means of structure sharing (again indicated by means of coindexing the attributes' values) between the attribute MOM\_SPEC, which is related to the behaviour of the figure, and the EVENT\_LAYERS | MANNER attribute.

Manner modifiers like *mit festem Schritt* 'with a firm step' in (2.d) and *auf Zehenspitzen* 'on tiptoe, tiptoeing' in (2.e) can also be treated as values of the MOM\_SPEC attribute, albeit ones with a more elaborate internal structure, due to their complex set-up as PPs. Both PPs contribute to the manner event layer and hence require appropriate structure sharing between the MOM\_SPEC and the MANNER attributes. I will not go into the details here and leave it at this short note.

### 2.3 VP level manner adverbs and dependency constraints

This chapter so far dealt with additional manner-of-motion characterizations and the expression of concurrent activities which are independent from the motion itself. In order to top off the considerations on mode-and-manner specifications expressed externally to the finite verb, I will now discuss adverbs like *schnell/quickly*, *fast* and *leise/quietly*, like in (2.b) and (2.c) above. The analysis of *schnell/quickly* and other velocity adverbs will induce us to establish an additional representational device, so-called dependency constraints, which will allow us to represent systematic dependencies between distinct but related values of different attributes in a frame.

Velocity adverbs like *schnell*, *langsam*, *quickly*, *fast*, *slowly* etc. can in principle be applied to any kind of activity. By their very nature, activities occur in time, and the adverbs of velocity in question express an indefinite measure on the time-related mode of execution of an activity by the actor; see (3.a). Adverbs like *leise*, *laut*, *quietly*, *loudly* – henceforth called “audibility adverbs” – are applicable to any activity whose execution may emit a sound. They express an indefinite measure on the acoustic properties of the execution of an activity by the actor; see (3.b):

- (3) a. walk/act/play/work/eat/drink/speak/write/think ...  
quickly/slowly  
b. walk/act/play/work/eat/drink/speak/write/?think ...  
quietly/loudly

When applied to activities like those in (2.a) and (2.b), I represent the meaning contribution of both velocity adverbs and audibility adverbs by a subattribute EXEC\_MODE (for “mode of execution”) of the agentive figure’s BEHAVIOUR attribute.<sup>5</sup> The two types of adverbs differ as follows: On the one hand, when a velocity adverb is applied to a motion verb, the value of this attribute affects the value of the path’s VEL(ocity) attribute profiled by the verb. In the case of *walk quickly/slowly*, the behavioural characterization as ‘quick’ or ‘slow’ causes the dimensional velocity attribute of the path to assume the values ‘high’ or ‘low’, respectively. In the case of audibility adverbs, on the other hand, there is no such impact on any property of the path and hence no dependency between the EXEC\_MODE attribute associated with the figure’s behaviour and the PATH attribute in the frame representation.

The kind of dependency between the values of different attributes in a frame which we observe for velocity adverbs is different from the ones that we have seen so far and that were captured by means of structure sharing between attributes, as represented by co-indexed attribute values. In the case of velocity adverbs, which in combination with a motion expression characterize a behaviour of the figure that has an impact on a dimensional property of the path, the values of the different attributes are not identical. Rather, they are distinct but systematically related. We capture their relatedness by means of an additional representational device in frames, namely so-called dependency constraints (DepCs). The notion of a DepC is inspired by Barsalou’s (1992) concept of contextual constraints. According to Barsalou (1992, p. 39), contextual constraints occur when one aspect of a situation constrains another, like the speed of a transportation which constrains its duration over a fixed time.

In the present case, I will introduce specific DepCs in order to express systematic dependencies between values of attributes in a frame. I will mark the determining element in a DepC by ‘ $\downarrow\textcircled{n}$ ’ and the determined element by ‘ $\uparrow\textcircled{n}$ ’. The index ‘n’ ranges over integers and serves as a unique identifier for

---

5 Note that both types of adverbs may also be applied to non-agentive processes where the subject has the role of a theme rather than an agent, like in *flow/drip/seep ... quickly/slowly/quietly/loudly*. It may be more appropriate to use an attribute like OCC\_MODE (for ‘mode of occurrence’), rather than EXEC\_MODE, in the case of thematic subjects.

<i>WALK</i>			
FIGURE	<b>BEHAVIOUR</b>	EXEC_MODE	↓@ <sub>1</sub> <b>quick</b>
EVENT_LAYERS	MANNER	walking_gait	
	PATH	ROUTE   ...	
		P_DIM   VEL	↑@ <sub>1</sub> <b>high</b>

Figure 6: The representation of *walk quickly*

<i>WALK</i>			
FIGURE	<b>BEHAVIOUR</b>	EXEC_MODE	<b>quiet</b>
EVENT_LAYERS	MANNER	walking_gait	
	PATH	...	

Figure 7: The representation of *walk quietly*

determining and determined elements in DepCs. The dependency will be indicated by the relation ‘↔’ between attribute structures.

The DepCs in (4) for the adverbs *quickly* and *slowly* convey that ‘quick’ and ‘slow’ as the corresponding values of EXEC\_MODE express certain energy levels in the execution of the activity that have an impact on the velocity with which the path is traversed in the course of the motion, indicated by the values ‘high’ and ‘low’, respectively:

- (4) a. DepC<sub>quickly</sub>: FIGURE | **BEHAVIOUR** | EXEC\_MODE | ↓@<sub>1</sub> **quick** ↔  
 EVENT\_LAYERS | PATH | P\_DIM | VEL | ↑@<sub>1</sub> **high**
- b. DepC<sub>slowly</sub>: FIGURE | **BEHAVIOUR** | EXEC\_MODE | ↓@<sub>2</sub> **slow** ↔  
 EVENT\_LAYERS | PATH | P\_DIM | VEL | ↑@<sub>2</sub> **low**

The crucial difference between velocity and audibility adverbs is shown in figures 6 and 7. For reasons of perspicuity, only the profiles of the adverbs are marked in boldface.

Figure 6 shows the impact of a velocity adverb like *quickly* on the dimensional properties of the path. This is accomplished by a DepC between the value of the EXEC\_MODE attribute of the figure, where the meaning contribution of the adverb is represented, and the VEL attribute of the path. By contrast, an acoustic quality adverb like *quietly* does not have any such impact. Figure 7 hence does not involve any DepC.

Figure 6 shows a dependency constraint between the BEHAVIOUR attribute of the figure and a subattribute of PATH, i.e., a dependency between dynamic elements of the motion event, namely the execution mode of an activity and the speed with which an evolving path is traversed. However, we also need to

provide for the possibility of dependencies between static (including, e.g., the figure's PROPERTY attribute), as well as between static and dynamic, elements of a situation. As a matter of fact, in section 3.6. I will propose to account for the connection between the intrinsic orientation of a ground object – i.e., a static, even persistent property of an entity – and the preferred interpretation of the vertical direction of movement (upward vs. downward), like in the case of *to climb* and its German counterparts, in terms of a dependency constraint.

## 2.4 Depictive manner predicates

The last type of manner modification that will be considered here is exemplified in (2.f) by the PP *in Pantoffeln* 'in slippers'. Other instances of this sort of modifier are *barfuß* 'barefooted' and *mit Schirm* 'with one's umbrella'. I call this sort of modifier a depictive predicate, following, *inter alia*, the nomenclature in Maienborn (1996, pp. 161 ff.) and Dolińska (2012, pp. 115 ff.) (who however discuss different instances of depictive predicates than the ones in focus here).

In the present approach, I treat depictive predicates as additional predications that pertain to the figure but are not immediately linked to the situation designated by the verb: After all, one cannot only walk, shuffle, sneak away or amble, but also read, cook, eat or play the piano in one's slippers or barefooted. This is different in the case of manner modifiers like *mit festem Schritt* 'with a firm step' and *auf Zehenspitzen* 'on tiptoe, tiptoeing' mentioned in section 2.2, which are related to the motion itself and are thus treated as values of the MOM\_SPEC attribute, which represents additional characterizations of the manner of motion (note that if you read, cook, eat or play the piano on your tiptoes, this refers to your posture or movement while you carry out the particular activities).

Depictive predicates like *in Pantoffeln* 'in slippers' and *barfuß* 'barefooted' will thus be represented under the PROPERTIES attribute of the figure, which was briefly introduced in section 2.2, figure 1. Being concerned here with the overall direction of the analysis rather than the subtleties of the choice of adequate frame attributes, I will content myself in the present context with the tentative representation in figure 8.

Similarly to figures 2 and 3, the fact that the property of wearing slippers holds for the time occupied by the main motion activity of *walking* is not represented separately by specific attribute structures. The concurrence again follows from the unique temporal interpretation of the entire frame and all of its components (which in the full-fledged version of the frame would be determined by the unique EVENT\_PROPERTIES attribute structure; cf. Herweg, 2020). Although the value of the PROPERTIES attribute complex in figure 8 does not have any explicit impact on other frame attributes, there may well be some

<i>WALK</i>				
FIGURE	<b>PROPERTIES</b>	<b>DRESS</b>	<b>FOOTWEAR</b>	<b>slippers</b>
EVENT_LAYERS	<b>MANNER</b>	<b>walking_gait</b>		
	<b>PATH</b>	...		

Figure 8. The representation of *walk in slippers*

valid commonsense assumptions about the mode of execution as well as about the manner and the velocity of the motion. So, the motion will probably be executed quietly and slowly when the moving entity wears slippers, and walking, shuffling, ambling or sneaking away are more likely manners than running, racing or rushing. But these are defeasible inferences at best, whose explanation has to be delegated to a theory of non-monotonic reasoning which is not in scope of the present study.

### 3 Varieties of manner of motion

In section 1 I already gave a brief overview of the types of manner verbs that will be addressed in the present study, namely:

1. verbs of type *S\_MOM*, which describe manners that are predominantly associated with the spatial dimension of a motion (e.g., *to walk*);
2. verbs of type *T\_MOM*, which describe manners that are distinctly temporal in nature (*to dash*, *to ramble*);
3. verbs of type *MC\_MOM*, which point to a mental constitution or state of mind that goes along with the motion (e.g., *to hasten*);
4. verbs of type *M\_MOM*, which describe a motion that is performed in a specific medium of localization (e.g., *to swim*);
5. verbs of type *I\_MOM*, which describe a motion that relies on an instrument (e.g., *to bike*);
6. and finally verbs of type *F\_MOM*, which describe manners that involve the exertion of a force against an object (e.g., *to climb*).

Each of these types defines which attributes and profiles are constitutive for the respective frame within or in addition to those of the general *LOCOMOTION* frame. We will see that a number of MoMVs show characteristics of more than one basic manner type. Formally, this is captured by assigning compound types  $T_1$  &  $T_2$  to the verbs in question.<sup>6</sup>

6 The underlying logic of verbal types is explained in Herweg (2020, 2021). In short, a complex type  $T_1$  &  $T_2$  adds the two profiles associated with  $T_1$  and  $T_2$ , including

I consider the proposed type-specific profiles on the basic locomotion frame first and foremost as an articulation of the potential that the verbs in question offer for the concrete interpretation of a motion construction in a given context. This means that contextual and situational factors may prevent this semantic potential from being used to full capacity in a given occurrence of a verb. In other words, depending on the context, individual components of the proposed frame representations may be more or less pronounced in the final interpretation of a motion construction. In this sense, the proposed profiled frames may well be understood as being prototypical. This applies in particular to those verbal types which add meaning elements that go beyond the primarily spatial aspects of motion events (i.e., direct specifications of the core attributes MANNER and PATH). The types *MC\_MOM* and *T\_MOM* are particularly noteworthy in this respect: Especially these types, which often appear as components in complex types  $T_1$  &  $T_2$ , can bring in their specific aspects of meaning more strongly (up to dominance) or more weakly (possibly present only in vestiges) in such type combinations, depending on the context of use.

The present chapter will only present some short sample analyses of the different types of MoMVs in order to give an idea of the general approach put forward here.<sup>7</sup> The proposed informal meaning descriptions and formal representations should be taken *cum grano salis*. A more in-depth study of this widespread field which goes beyond a mere collection of simplifying illustrations has to be deferred to separate work.

### 3.1 Space-dominant manner of motion: the type *S\_MOM*

The type *S\_MOM*, which is short for ‘space-dominant manner of motion’, is intended to cover motion activities whose characteristics can to a large extent be described in terms of spatial configurations of (parts of) the figure in relation to (parts of) the ground. Examples of this type of motion are the basic human gaits of walking and running. Wikipedia (en.wikipedia.org/wiki/Walking, 2023/3/1) summarizes the standard physiological distinction between these gaits (see also Diersch, 1972, pp. 65, 135 for similar definitions) in terms of the contact between the feet of the figure and the ground:

---

their corresponding attributes, to the frame. The denotation of a frame of type  $T_1$  &  $T_2$  can be thought of as the intersection of the two sets of situations described individually by  $T_1$  and  $T_2$ .

7 Other manner classifications have been proposed in the literature, such as the one in Slobin et al. (2014). See Cifuentes Ferez (2007) for an overview of other classification systems; cf. also, e.g., the classification of Thai manner-of-motion verbs in Takahashi (1997). I consider the manner categories suggested in these works to be subtypes of the general types put forward here.



Walking requires that there is always one foot in contact with the ground and that there is a short period of double-support, i.e., a period where the body is supported by both feet. By contrast, running means that the body is supported only by one foot at a time and that there is a short period where both feet are off the ground (i.e., an aerial phase).

Mani & Pustejovsky (2012, p. 105), in their seminal work towards a logical foundation of the semantics of motion expressions, suggest that quite a few manner-of-motion verbs can be differentiated on the basis of topological and mereotopological constraints concerning the figure and the ground during the movement. As a matter of fact, the examples they discuss are mostly verbs that are treated as predicates of type *S\_MOM* in the present approach. Mani & Pustejovsky use notions from RCC-8, the Region Connection Calculus (Randell et al., 1992), a calculus which basically extends Allen's (1983) temporal Interval Calculus into two-dimensional space in order to facilitate representing, as well as reasoning about, qualitative spatial relations. Employing the RCC-8 relations of two regions being externally connected or disconnected, Mani & Pustejovsky characterize the basic gaits of walking and running by the fact that throughout the movement there is a continuous alternation between a state where a proper part of the figure is externally connected with the ground and a state where this part of the figure and the ground are disconnected. By contrast, hopping and bouncing imply that the entire figure, rather than just one of its proper parts, alternates between being externally connected with and being disconnected from the ground. Manners such as sliding and rolling are characterized by the fact that there is an uninterrupted external connection between figure and ground.

A straightforward way to incorporate elements of the (mereo-)topological analysis into the present framework amounts largely to a direct translation of the Mani & Pustejovsky approach into the account of dynamic verbal predicates in Herweg (2021), which utilizes a variant of the Dynamic Interval Temporal Logic of Pustejovsky & Moszkowicz (2011) and Mani & Pustejovsky (2012). This means that the logical formalism that underlies the dynamic interpretation of the relevant verbal predicates in the FAMEu *LOCOMOTION* frame would be enhanced with (mereo-)topological concepts along the lines of those employed by Mani & Pustejovsky (2012), yielding frames of verbal types *RUN*, *WALK*, *BOUNCE*, *HOP*, *SLIDE*, *ROLL* etc. (for an outline of the general approach see Herweg, 2021).

Mani's & Pustejovsky's approach can, however, be considered only as a first approximation of a differentiation of manners of (our) type *S\_MOM*, for the following reasons: First, further refinement of the characteristic sequence of (mereo-)topological relations is necessary in order to distinguish, e.g., walking and running, as informally described above, as well as other gaits. This also requires further specification of the involved parts of the figure's body and

<i>LIMP</i>			
FIGURE	BEHAVIOUR	EXEC_MODE	laborious & unbalanced
EVENT_LAYERS	MANNER	walking_gait	
	PATH	...	

Figure 9. The representation of *to limp*

a much more differentiated view of their relations to the ground in order to express, e.g., how human running differs from *to hop*, *to bounce*, *to skip* etc. – all these manners include a short period when figure and ground are disconnected – in that running involves an alternation between the feet which are connected and disconnected with the ground. Second, it is doubtful that the approach carries over to manners such as striding and limping, i.e., specific subgaits of walking which involve some peculiarity or even impediment, features which apparently cannot be primarily captured in terms of temporal sequences of (mereo-)topological relations.

In conclusion, even though (mereo-)topological concepts, embedded in a dynamic logic, provide a compelling formal framework for the semantic analysis of MoMVs, the specific gaits which cannot be captured with these means alone rather call for enriching the basic *LOCOMOTION* frame with additional attributes, so that we eventually arrive at a mixture of a dynamic plus (mereo-)topological foundation of verbal predicate types and the use of explicit attributes in frames. As an empirically well-studied example illustrating this approach, consider the manner-of-motion verb *to limp* and its German counterpart *hinken*. In her contrastive corpus-based in-depth study of English and German verbs, Snell-Hornby (1983, pp. 137 ff.) assigns to these verbs the class ‘clumsy, unsteady motion’, which covers deviant or impeded modes of walking that are conditioned by the physical or mental state of the figure. More specifically, according to this study, *to limp* and *hinken* describe an irregular and labourious walking caused by lameness which can even indicate a permanent disability.

The partial frame in figure 9 serves as a rough approximation of the representation of *to limp/hinken* in order to illustrate the general approach. The frame’s overall type is the verbal predicate *LIMP*, which designates a specific subgait of walking and of the general type *S\_MOM*. This predicate puts additional constraints on the basic walking gait, for whose representation we again make use of the “mode of execution” attribute *EXEC\_MODE* already employed in section 2.3. The profile of the verb thus expands the figure’s attribute structure accordingly, as indicated in figure 9. The value ‘laborious & unbalanced’ of *EXEC\_MODE* should simply be considered to be short for Snell-Hornby-like feature complexes and just serves to illustrate the general approach.

As a final example, which shows that the type *S\_MOM* may be conflated with other manner types such as a manner related to the figure's state of mind (which is covered by our type *MC\_MOM*, see below, section 3.3), consider the verb *to stride* and its close German relative *schreiten*. Snell-Hornby (1983, pp. 135 f.) assigns to both verbs the class 'measured, labourious motion'. More specifically, Snell-Hornby describes striding as an energetic and purposeful mode of walking, executed with long, swift steps, whereas the German verb *schreiten* focuses on conscious dignity in walking, with regular and measured steps, an upright posture and a controlled execution, but without affectation. Similarly, in her in-depth study of German motion verbs, Diersch (1972, pp. 122 ff.) characterizes the meaning of *schreiten* as a subgait of walking with a controlled body, an upright posture, an even step rate, plus an at the same time elated and controlled state of mind.

In the present approach, the manners associated with *to stride* and *schreiten* will be assigned the compound manner type *S\_MOM & MC\_MOM*. The verbs' profiles again augment the FIGURE attribute in a specific way that will be briefly dealt with in section 3.3, which elaborates on the type *MC\_MOM* and will show that manifestations of the figure's state of mind may play a significant role in manners of motion (as also witnessed by Zlatev et al., 2021).

### 3.2 Time-dominant manner of motion: the type *T\_MOM*

The type *T\_MOM*, which is short for 'time-dominant manner of motion', is intended to cover motion activities whose primary characteristics are temporal in nature. Examples are activities designated by verbs like *to race*, *to dash*, *to sweep*, which refer to movements that are (judged to be) particularly fast, and verbs like *to ramble*, *to stroll*, *to dander*, *to saunter*, which rather refer to movements that are (judged to be) slow. Being movements, the designated activities of course still have a spatial dimension, but their distinctive features are related to time rather than space. In contrast to verbs of type *S\_MOM*, the characteristics of verbs of type *T\_MOM* cannot – at least not significantly – be described in terms of spatial configurations of (parts of) the figure in relation to (parts of) the ground. Some of the verbs are not at all associated with any particular gait (*to rush*, *to dash*, *to sweep*), while others may be weakly associated with a specific bodily behaviour (*to ramble*, *to dander*, also *to amble*). However, I consider these associations to be highly indeterminate and subject to a great deal of situational variation and therefore do not include them in the primary information these verbs convey. All the verbs in question rather highlight the amount of time that the agent spends or intends to spend on moving from one place to another, up to a point where any notion of changing locations is almost completely obscured by time-related aspects.

<i>DASH</i>			
FIGURE	BEHAVIOUR	EXEC_MODE	↓ <sub>⊙</sub> <sub>1</sub> high_rate
EVENT_LAYERS	MANNER	dashing	
	PATH	ROUTE   ...	
		P_DIM   VEL	↑ <sub>⊙</sub> <sub>1</sub> high

Figure 10. The representation of *to dash*

<i>SAUNTER</i>			
FIGURE	BEHAVIOUR	EXEC_MODE	↓ <sub>⊙</sub> <sub>1</sub> leisurely
EVENT_LAYERS	MANNER	sauntering	
	PATH	ROUTE   ...	
		P_DIM   VEL	↑ <sub>⊙</sub> <sub>1</sub> low

Figure 11. The representation of *to saunter*

The profiles that we propose for verbs of type *T\_MOM* in figures 10 and 11 again make use of the ‘mode of execution’ attribute EXEC\_MODE in the figure’s BEHAVIOUR structure. The frame also employs Dependency Constraints (DepCs), as introduced in section 2.3, in order to represent the impact of the value of the figure’s EXEC\_MODE on the value of the VEL(ocity) attribute in the path’s dimensional representation. The values of EXEC\_MODE are borrowed from some vivid informal meaning characterizations that can be found in Snell-Hornby (1983, pp. 133, 142) and Cifuentes Ferez (2007, p. 120). They only serve to illustrate the general approach, which also holds for the value of the MANNER attribute, whose specifics are not relevant in the present context. The sample frames are just intended to show that the verbs’ profiles are still those of manner-of-motion verbs; the manner attribute is therefore not suppressed but remains silent about any specific gait.

The (partial) frames in figures 10 and 11 exhibit two characteristic principles of the profiling mechanism: The basic FIGURE attribute of the *LOCO-MOTION* frame is augmented with a complex BEHAVIOUR attribute, and the dependency constraint establishes a relation between separate substructures of the frame (cf. section 1).

Even more pronounced than in the case of verbs of type *T\_MOM*, many verbs of type *S\_MOM* exhibit a distinct intentional, emotional or attitudinal element. Meaning components that refer to characteristics of the state of mind of the moving entity have been pointed out notably by Diersch (1972) and Snell-Hornby (1983) in their corpus-based studies of verb meanings. Diersch (1972, pp. 130 ff.) cites evidence from text collections that German *schlendern* ‘saunter, stroll, amble’ is often associated with a placid, calm and purposeless state-of-mind of the agent, whereas German *eilen* ‘rush, hurry’ (Diersch, 1972,

pp. 151 ff.) usually comes with an active and exerted attitude. Similar results are reported in Snell-Hornby (1983, p. 133) for English *to amble*, *to saunter*, *to stroll* etc., as well as for German *schlendern*, which are in general associated with a leisurely, easy-going attitude of the agent, whereas a verb like *to rush* expresses the agent's inner impetus of urgency (Snell-Hornby, 1983, p. 144). In a similar vein, Cifuentes Ferez (2007) characterizes *to roam* and *to saunter* as aimless motion without any special purpose.

Observations like these lead us to the next type of manners of motion, namely those which exhibit a strong intentional, emotional or attitudinal component.

### 3.3 Mental-constitution – dominant manner of motion: the type *MC\_MOM*

The type *MC\_MOM* covers motion verbs that suggest a pronounced mental constitution or state of mind of the moving figure which represents the figure's intentions, emotions, attitudes etc. Some examples were already given in the previous sections. In the present section I will focus on two German verbs which have been extensively studied by Diersch (1972), namely *eilen* 'rush, hurry' and *hasten* 'hasten, hustle'. These verbs are semantically quite close but reveal some interesting differences. In both cases, the described motions are independent from any specific gait and can be performed in different ways, e.g., by walking, running, driving, biking etc. Both verbs primarily express that the agent strives for a fast motion in order to quickly pass some distance, like in the following examples:

- (5) a. Er eilte von der Bibliothek über den Korridor in den Hörsaal.  
'he rushed / hurried from the library across the corridor into the lecture hall'
- b. Er hastete aus dem Büro durch die Altstadt zum Theater.  
'he hustled / hastened out of the office through the old town to the theatre'

There is, however, an important difference: While *eilen* is – apart from the agent's endeavour for speed – neutral with regard to the state of mind of the agent, *hasten* attributes to the agent a tense, driven and agitated state of mind (cf. Diersch, 1972, p. 156). This difference can be tracked down in the contexts in which these verbs typically appear: According to Diersch's corpus studies (cf. Diersch, 1972, pp. 151 f.), *eilen* appears equally with characterizations of the velocity of motion (e.g., *so schnell wie möglich* 'as quickly as possible'), of the figure's gait and posture (e.g., *mit langen Schritten* 'with long steps'), of accompanying activities (e.g., *flüchtend* 'fleeing') as well as

<i>EILEN</i>				
FIGURE	PROPERTIES	SOM	INT	perform_quick_motion
EVENT_LAYERS	MANNER	eilend		
	PATH	...		

Figure 12. The representation of German *eilen* ‘rush, hurry’

<i>HASTEN</i>				
FIGURE	PROPERTIES	SOM	INT	perform_quick_motion
EVENT_LAYERS	MANNER	hastend	EMO	agitated
	PATH	...		

Figure 13. The representation of German *hasten* ‘hasten, hustle’

his / her mood, which can be positive or negative (e.g., *zielbewusst* ‘purposeful’, *erschreckt* ‘appalled’). In contrast, *hasten* predominantly appears with descriptions of the figure’s mood – which is mostly negative (e.g., *wütend* ‘angry’) – and accompanying activities, which mostly indicate a substantial effort (e.g., *schnaufend* ‘panting’). However, specifications of velocity, gait and posture are rare (cf. Diersch, 1972, p. 155).<sup>8</sup>

In the representations shown in figure 12 and figure 13, the profiles associated with these verbs expand the FIGURE attribute with state-of-mind related properties (attribute SOM), such as the figure’s intention (INT) and – in the case of *hasten* – his / her emotion or affect (EMO).<sup>9</sup> The representations show that *hasten* essentially adds a particular state of mind to the intention to perform a motion with high speed.

Neither representation explicitly refers to the VEL(ocity) of the motion because the verbs seem to focus more on the (ascribed) intention of the agent than on the actual execution of the motion (which actually does not have to

8 Queries by the author in the German corpora ‘DeReKo’ and ‘Wortschatz Leipzig’ [www.ids-mannheim.de/digspra/kl/projekte/korpora, www.wortschatz.uni-leipzig.de/de; accessed November 2022] produced a substantial (though not yet exactly quantified) number of cases where *hasten* occurred with final clauses (e.g., *um den Zug zu erreichen* ‘in order to catch the train’) and PPs like *auf dem Weg zur Arbeit* ‘on the way to work’ or *in panischer Angst* ‘petrified’ (lit.: *in panic fear*). These adverbial modifications can be understood as state-of-mind related, often intentional patterns of explanation.

9 I would like to emphasise again that the proposed representations are only meant to show the general direction of the analyses and that the designations of intentions, emotions etc. (i.e., the values of the corresponding attributes) are only for illustration. In a further stage of elaboration, these designations would have to be underpinned by corresponding psychological theories, such as Ekman’s (1992) theory of emotions (or its later enhancements; cf. Hofmann et al., 2020).

be particularly fast). If corroborated by additional evidence, the value of the path's VEL attribute could be adjusted accordingly by means of a dependency constraint between the figure's intention value and the path's velocity value.<sup>10</sup>

It appears that in the majority of cases the type *MC\_MOM* is not the sole manner type associated with a verb but rather serves as an additional condition on other types like *T\_MOM* or *S\_MOM*. The verbs in question can thus be assigned complex types such as *T\_MOM & MC\_MOM* (e.g., *to rush*, *to hurry*; *eilen*) or *S\_MOM & MC\_MOM* (see section 3.1 on *to stride* and German *schreiten*).

### 3.4 Medium-related manner of motion: the type *M\_MOM*

The type *M\_MOM* serves to represent those kinds of motion that are performed in a specific medium of localization (cf. Kaufmann, 1995), as in the case of *to swim* and *to fly*. The medium of localization can be identical with or directly be determined by the explicit ground object, like in *to fly through the air* and *to swim through the English Channel*, resp., or it can be semantically detached from the ground object, like in *to fly across the English Channel* or *to swim from Cap Gris-Nez to Dover*, where the medium needs to be established by additional inference (neither is the English Channel an airspace, nor are the sites Cap Gris-Nez or Dover waterways). We will hence make use of a separate attribute MED for 'medium of localization' that is independent from the ground. The profiles of the verbs in question expand the figure's PROPERTIES attribute (see above, figure 1 and section 2.4) with a localization property LOC which specifies a relation L\_REL – the same localization relation that is used in the attribute structures of the different route segments – between the figure and the medium. In the case of *to swim* in figure 6, the type of the medium is 'liquid'. In the case of *to fly*, the type of the medium is 'air'. The spatial relations 'in' and 'at' are deliberately left vague and are intended to cover the ranges of interpretations of the corresponding prepositions. The

---

10 If at all, I consider such a relation between intention and velocity to be more appropriate for *eilen* than for *hasten*. However, I would prefer to assume a weak constraint at best – a relation for which I propose the concept of weighted dependency constraints in Herweg (forthcoming). It is instructive for the study of these two verbs to look at the corresponding nouns. DUDEN describes their meanings as follows: (a) *Eile*: *Bestreben, Gedrängtsein, etwas so schnell wie möglich zu erledigen* ['hurry, rush: striving, being in a push to get something done as quickly as possible']; (b) *Hast*: *große, überstürzte Eile; (oft von innerer Unruhe oder der Angst, nicht rechtzeitig fertig zu werden, verursachtes) planloses, aufgeregtes Handeln* ['haste: great, hurried rush; (often caused by inner restlessness or fear of not finishing in time) haphazard, agitated action'] ([www.duden.de/rechtschreibung/Hast#bedeutung](http://www.duden.de/rechtschreibung/Hast#bedeutung), accessed on March 5th, 2023; translations by [www.deepl.com/translate](http://www.deepl.com/translate)). Both definitions emphasize the agent's state of mind and do not address the actual velocity of the motion.

<i>SWIM</i>					
FIGURE	PROPERTIES	LOC	L_REL	in	
			MEDIUM	LIQUID	
EVENT_LAYERS	MANNER	swimming			
	PATH	ROUTE	INIT	L_REL	at
				GROUND	cap_gris_nez
			INTER	L_REL	LOC_REL
				GROUND	PHYSICAL_OBJECT
			FIN	L_REL	at
				GROUND	dover

Figure 14. The representation of *to swim from Cap Gris-Nez to Dover*

representation in figure 14 shows that the medium and the ground object(s) are independent from each other.

A verb of type *M\_MOM* like *to swim* thus profiles a complex attribute structure according to which the figure performs a specific motion activity in a medium of a particular sort. The manner in which the motion is executed – in the present case represented by the general type ‘swimming’ – comprises a family of bodily activities which could be described more specifically as crawl, backstroke, breaststroke, dolphin butterfly stroke or (at least in the case of German *schwimmen*) even simple floating. These activities involve particular spatial configurations of (parts of) the figure in relation to the medium, so that we should assume for the verbs in question also a discernible spatial manner component, which can be captured by the compound type *M\_MOM & S\_MOM*.

### 3.5 Instrument-related manner of motion: the type *I\_MOM*

The type *I\_MOM* comprises manners of motion whose execution relies on a separate device, in addition to the human body, which serves as an instrument in the motion activity. This includes activities such as cycling, rowing, sailing, skating, skiing, sledding etc. (cf. Kaufmann, 1995; Pourcel, 2004). For a simple representation of this manner type we can work with the frame inventory that is already available and again utilize the EXEC\_MODE attribute in the figure’s BEHAVIOUR structure, like we did for the types *T\_MOM* and *S\_MOM*. In the proposed frame, verbs of type *I\_MOM* augment the EXEC\_MODE with an instrument whose sort is represented by the type of the value of the associated attribute INSTR. Figure 15 shows the sample profile of the verb *to cycle*.

Similar to the type *M\_MOM*, verbs of type *I\_MOM* profile frame structures which reflect that the figure performs a specific motion activity in relation to a separate entity – here: the instrument – which is an intrinsic ingredient of the respective manner of motion. The manner in which the motion is executed consists in specific bodily activities like stepping onto the pedals, pulling and



<i>CYCLE</i>			
FIGURE	BEHAVIOUR	EXEC_MODE	employ_instrument
			INSTR   <i>BICYCLE</i>
EVENT_LAYERS	MANNER   cycling		
	PATH   ROUTE   ...		

Figure 15. The representation of *to cycle*

pushing the oar blades, navigating a sailing boat etc. These activities again involve a spatial manner component which calls for assigning to the verbs in question the compound type *M\_MOM* & *S\_MOM*, plus possibly the type described in the next section.

### 3.6 Force-exerting manner of motion: the type *F\_MOM*

The manner type *F\_MOM*, which is short for ‘force-dominant manner of motion’, is used for verbs which designate a motion that involves the exertion of a force against an overt or tacit object. This object serves as an antagonist in the activity, like in vertical movements as designated by *to climb*, as well as by German *klettern* and (the manner-of-motion reading of) *steigen*<sup>11</sup> (cf. Geuder & Weisgerber, 2008). In the account put forward by Geuder & Weisgerber (2008), whose core elements are adopted in the present study, these verbs of vertical movement describe a manner of motion that specifies force configurations on a path, namely the manner ‘upward force exertion (against the ground)’. This manner implies a direction (‘upward’), which, however, applies only to the force exerted against a ground object, such as (the steps of) a ladder. The examples (6.a–c) show that the overall direction of the motion is in principle independent of this force-related direction and can be upward as well as downward. Only with certain types of ground objects does a preferred interpretation of direction emerge, such as downward in (6.d) and upward in (6.e):

- (6) a. to climb up / down a mountain  
 b. auf das / vom Dach klettern ‘to climb onto / from the roof’  
 c. in ein Auto steigen ‘to climb into a car’  
 d. in einen Schacht steigen/klettern ‘to climb into a manhole’  
 e. in eine Baumkrone steigen/klettern ‘to climb into a treetop’

11 The restriction to the manner-of-motion sense of *steigen* excludes the pure directional reading of *steigen* like in *Der Ballon steigt* ‘the balloon rises’, which involves reference to a freely suspended entity and excludes downward movement (cf. Gamerschlag et al., 2014).

<i>STEIGEN_MM</i>									
FIGURE	LEGS	①							
	BEHAVIOUR	EXEC_MODE	FORCE	AGONIST	①				
				ANTAGONIST	②				
				EXERTION	DIR	upwards			
				INTENSITY	medium_high				
EVENT_LAYERS	MANNER	③ stepwise_gait							
	PATH	ROUTE	INIT	L_REL	LOC_REL				
					GROUND	PHYSICAL_OBJECT			
			INTER	L_REL	~in				
				GROUND	④ manhole	STEPS	②		
					OS ↓ ⑤				
			FIN	L_REL	in				
				GROUND	④				
		P_DIM	DIR	vertical	↑ ⑥				

Figure 16. Frame for the force expression *in einen Schacht steigen* ‘climb into a manhole’

Figure 16, which shows the FAMEu frame for *in einen Schacht steigen* ‘to climb into a manhole’, incorporates the crucial elements of the frame-semantic analysis of the manner-of-motion reading of German *steigen* ‘to climb’ – henceforth represented by the verbal type *STEIGEN\_MM* – as suggested in Gamerschlag et al. (2014), which in turn elaborates on the ideas in Geuder & Weisgerber (2008). In terms of the present approach, the profile associated with *STEIGEN\_MM* expands the figure’s BEHAVIOUR attribute with the representation of a specific force configuration. The general picture is that the exertion of a force effectuates the execution of a particular manner of motion that causes a change of location. In the present example, the manner is executed with medium-to-high intensity in a stepwise fashion. The figure’s legs constitute the agonist of the force, whose antagonist is provided by a suitable object associated with the manhole. Just for the purpose of illustration we assume that some steps attached to the manhole form the antagonist.

Over and above the approximate translation of the account in Gamerschlag et al. (2014) into the present framework, the proposal in figure 16 incorporates a frame attribute OS for object schemes. Object schemes in the sense of Lang (1989) cover the dimensional properties of physical objects and primarily serve to determine which dimension of an object counts as its height, width, length, depth or thickness in a given context. The idea in the case at hand is that the verb *steigen* is neutral with regard to upward or downward movement – hence P\_DIM | DIR | *vertical* by virtue of its profile – and that the orientation of the ground object may further constrain the interpretation of the direction of the movement. The vertical orientation of an object can be intrinsically determined (downward in the case of manholes, canyons etc.; upward in the case of towers, treetops, helicopters etc.), or it can very much depend

on the particular situation (think of climbing into a cave, a tube or a pipe; see also [6.c]). In order to account for the fact that the orientation of an object in many cases needs to be calculated from its OS and the circumstances of the respective situation, I propose to represent the impact of the orientation of the ground object in a particular situation on the interpretation of the overall direction of the motion – here: the climbing – again by a dependency constraint (cf. section 2.3). This  $\text{DepC}_{\text{orient}}$  (which I can only hint at here and whose details I will have to leave open for reasons of complexity) further constrains the value of the path's direction, which the verb only (under-)determines as *vertical*, to a more specific upward or downward verticality. Amended with a suitable account of defeasible interpretation preferences (which again will be subject to a separate study on weight assignments in frames; see Herweg, forthcoming), this device is designed to capture the preferred interpretation of the direction of the motion, i.e., a downward motion in (6.d) *in einen Schacht steigen/klettern* 'to climb into a manhole', as opposed to an upward motion in (6.e) *in eine Baumkrone steigen/klettern* 'climb into a treetop'.

Since manner-of-motion verbs of vertical movement expand both the MANNER and the PATH attribute (plus the figure's BEHAVIOUR attribute), they concurrently profile the two event layers manner and path, rather than just a single event layer.

It needs further consideration to determine whether or not all locomotion verbs are uniformly assigned a FORCE attribute, or if there rather is a clear distinction between force-related and non-force-related verbs. Some strong candidates for force-related verbs were already mentioned in the previous section on verbs of type *L\_MOM*. In case of a pervasive FORCE attribute, the concept of weighted frame elements may again be useful in order to differentiate among verbs in terms of the weights assigned to the FORCE attribute, depending on the role that forces play in their semantics. As mentioned before, this topic will be revisited in a forthcoming study (Herweg, forthcoming).

#### 4 By way of conclusion: Cognitive verb clusters as evidence for the microstructure of the MANNER attribute

This article put forward some additional pieces of equipment that will be beneficial on our journey towards a comprehensive frame-semantic account of motion expressions in typologically different languages, the principles of which have been outlined in Herweg (2020, 2021). Our point of departure is the FAMEu *LOCOMOTION* frame, which provides an expressive structure that allows to represent major elements of the linguistic description of motion events. Most importantly, the *LOCOMOTION* frame features elaborate

representations of the two fundamental event layers that are effective in motion descriptions, namely the manner layer and the path layer.

The crucial theoretical concept used for a fine-grained differentiation among the semantics of intransitive motion verbs is the characteristic profile that a motion verb imposes on the *LOCOMOTION* frame. The particular profiles of motion verbs may select, deselect and augment substructures of the general frame, as well as elaborate on significant relations between different substructures. We saw instances of each of these mechanisms at work in the suggested differentiations in the multi-faceted expression of manner of motion.

The FAMEu approach is intended to be applied not only to the semantic analysis of motion verbs, but also to modelling some of the cognitive aspects of the interpretation of motion expressions, as identified in psycholinguistic research. And there are indeed further empirical observations in favour of some of the proposed distinctions from a pilot experiment that was targeted on the identification of semantic factors that determine preferences in the aspectual class interpretation of motion expressions in German (cf. Herweg & Gerwien, 2016). In order to isolate the contribution of verb meanings to aspectual class preferences, subjects were presented with motion descriptions in which the linguistic context of the motion verb had been reduced to a minimum, namely a directional goal PP (*in* + NP[acc]), which creates a telic context, and its locative variant (*in* + NP[dat]), which creates an atelic context. Subjects were asked to judge the acceptability of about 80 German motion verbs complemented by either a directional or a locative PP on a continuous scale supplied via a slider on a computer screen. The expected outcome was that subjects would produce acceptability judgements that provide a clue to identify verbs with different degrees of preference for telic and atelic interpretations.

The acceptability judgements produced different clusters of verbs. A notable result in the context of the present study is the fact that the German equivalents of the above mentioned verbs with a predominantly space-related manner (verbs of type *S\_MOM*, in terms of the FAMEu type system) showed no discernible preference for either a telic (directional) or atelic (locative) context, whereas the German equivalents of the verbs that designate a predominantly time-related manner (type *T\_MOM*) with a high velocity value showed a significant preference for a telic (directional) context; atelic (locative) contexts were more or less completely rejected.

Quite interestingly, the experiment also revealed a group of verbs with remarkably low ratings in telic directional contexts, namely verbs like *bummeln* ‘amble, saunter’, *flanieren* ‘dander, stroll’, *streunen* ‘stray’ and *stromern* ‘roam’. While these findings can to some extent be explained by the aspectual class of the verbs – there is good reason to assume that the latter two verbs are lexically atelic (cf. Maienborn, 1990; Herweg, 2020) –, this does not account for the

entire group in question, as the former two verbs' aspectual class is lexically underspecified, rather than atelic.<sup>12</sup> A promising move towards an explanation of these judgements, on top of considerations about aspectual classes, could be built upon the system of manner types proposed in the present paper. I hence conjecture that the verbs of this group share relevant semantic properties with several of the manner-of-motion verb types introduced in section 3: First, they inherit features from the predominantly time-related type *T\_MOM*, on whose P\_DIM | VEL frame component their profile imposes a low velocity value. At the same time, the verbs in question still exhibit some characteristics of the space-dominant type *S\_MOM*, as they also evoke a certain impression of a characteristic gait, although this is not their predominant interpretation. And finally, the verbs convey the idea of an aimless motion that primarily serves as a pastime, rather than a movement firmly aiming at a given goal – a trait which can be modelled utilizing the *MC\_MOM* type and verb-specific profiles imposed on the corresponding intention attribute. This conflation of particular type characteristics might give us a clue why these verbs do not completely resist, but strongly disfavour a decidedly telic context.

I have to leave it open to future experimental research to further investigate the principles that guide the cognitive construal of motion event representations and to shed more light on the interaction of (more or less) stable lexical semantics with defeasible standard assumptions and interpretation preferences.

## Acknowledgments

I am grateful to Johannes Gerwien and Kristian Nicolaisen for many helpful comments on an earlier version of this article.

## References

- Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11), 832–843.
- Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In A. Lehrer, & E. F. Kittay (Eds.), *Frames, Fields, and Contrasts. New Essays in Semantic and Lexical Organization* (pp. 21–74). Lawrence Erlbaum Associates.
- Carpenter, R. L. (1992). *The Logic of Typed Feature Structures: With Applications to Unification Grammars, Logic Programs and Constraint Resolution* (Cambridge Tracts in Theoretical Computer Science). Cambridge University Press.

---

12 Herweg (forthcoming) cites some examples from German corpora that corroborate this claim.

- Cifuentes Ferez, P. (2007). Human Locomotion Verbs in English and Spanish. *International journal of English studies*, 7(1), 117–136.
- Diersch, H. (1972). *Verben der Fortbewegung in der deutschen Sprache der Gegenwart. Eine Untersuchung zu syntagmatischen und paradigmatischen Beziehungen des Wortinhalts*. Akademie-Verlag.
- Dolińska, J. (2012). *Zur Klassifizierung der Prädikative* [Doctoral Thesis, Friedrich-Schiller-Universität Jena]. d-nb.info/1022011448/34
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion* 6(3–4), 169–200.
- Gamerschlag, T., Geuder, W., & Petersen, W. (2014). Glück auf, der Steiger kommt: A frame account of extensional and intensional steigen. In D. Gerland, C. Horn, A. Latrouite, & A. Ortman (Eds.), *Meaning and Grammar of Nouns and Verbs*. (pp. 115–144). Düsseldorf University Press. [urn:nbn:de:hbz:061-20140414-160807-7]
- Gerwien, J., & von Stutterheim, C. (2016, March 2). *Grammatical constraints on event packaging and potential effects on the segmentation of the perceptual stream. Motion events in language and cognition under a cross-linguistic perspective* [oral presentation]. Pre-CUNY workshop: Events in Language and Cognition – 2016, Gainesville, FL, United States.
- Gerwien, J., & Herweg, M. (2017). Aspectual Class (Under-)Specification in the Generation of Motion Event Representations – A Project Outline. *HUPLC – Heidelberg Papers on Language and Cognition 2017 (1)*. <https://dx.doi.org/10.11588/huplc.2017.0.37820>, URN (PDF): <https://nbn-resolving.org/urn:nbn:de:bsz:16-huplc-378209>
- Geuder, W., & Weisgerber, M. (2008, March 28). *Manner of Movement and the Conceptualisation of Force* (presentation). Journée d'étude „Il y a manière et manière“, Université d'Artois, Arras, France. [semarch.linguistics.fas.nyu.edu/Archive/Tk5YmEwN/MannerMovement\_slideswithanimations.pdf]
- Hendriks, H., Hickmann, M., & Pastorini-Campos, C. (2021). Running or crossing? Children's expression of voluntary motion in English, German, and French. *Journal of Child Language* (2021), 1–24. <https://doi.org/10.1017/S0305000921000271>
- Herweg, M. (2014). Spatio-temporal modification and the determination of aspect: a phase-theoretical account. In D. Gerland, C. Horn, A. Latrouite, & A. Ortman (Eds.), *Meaning and Grammar of Nouns and Verbs* (pp. 185–222). Düsseldorf University Press. [urn:nbn:de:hbz:061-20140414-160807-7]
- Herweg, M. (2020). Motion verb constructions, frames, and profiling. *Questions and Answers in Linguistics* 6(1), 1–30. [https://www.researchgate.net/publication/347326018\\_Motion\\_verb\\_constructions\\_frames\\_and\\_profiling#fullTextFileContent](https://www.researchgate.net/publication/347326018_Motion_verb_constructions_frames_and_profiling#fullTextFileContent)
- Herweg, M. (2021). Dynamic Event Types in Frame Semantics. The Representation of Change in FAMEu. Working Paper. *HUPLC – Heidelberg University*

- Papers on Language and Cognition* 2(1), 15–44. <https://doi.org/10.11588/huplc.2021.1.82301>
- Herweg, M. (forthcoming). *Towards modelling interpretation preferences in frame semantics. Weight assignment and propagation in FAMEu*. Draft version on Research Gate: [https://www.researchgate.net/publication/325403274\\_Towards\\_Modelling\\_Interpretation\\_Preferences\\_in\\_Frame\\_Semantics\\_Weight\\_assignment\\_and\\_propagation\\_in\\_FAMEudraft\\_v4\\_April\\_12th\\_2024](https://www.researchgate.net/publication/325403274_Towards_Modelling_Interpretation_Preferences_in_Frame_Semantics_Weight_assignment_and_propagation_in_FAMEudraft_v4_April_12th_2024)
- Herweg, M., & Gerwien, J. (2016, July 14). *Modeling Cognitive Mechanisms in Motion Event Construal: Aspectual (Under-)Specification* (Poster presentation). Heidelberg University Language & Cognition (HULC) Lab Summer Symposium 2016. Heidelberg, Germany.
- Hofmann, J., Troiano, E., Sassenberg, K., & Klinger, R. (2020). Appraisal theories for emotion classification in text. *Proceedings of the 28th International Conference on Computational Linguistics, Barcelona*, 125–138.
- Kaufmann, I. (1995). *Konzeptuelle Grundlagen semantischer Dekompositionsstrukturen. Die Kombinatorik lokaler Verben und prädikativer Komplemente*. Niemeyer.
- Lang, E. (1989). The semantics of dimensional designation of spatial objects. In M. Bierwisch, & E. Lang (Eds.), *Dimensional Adjectives: Grammatical Structure and Conceptual Interpretation* (pp. 263–417). Springer-Verlag.
- Langacker, R. W. (1983). *Foundations of Cognitive Grammar I: Orientation & II: Semantic Structure* (reproduced by L.A.U.T). Linguistic Agency of Trier.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites* (Vol. 1). Stanford University Press.
- Langacker, R. W. (2013). *Essentials of Cognitive Grammar*. Oxford University Press.
- Maienborn, C. (1990). *Position und Bewegung: Zur Semantik lokaler Verben* [IWBS-Report Nr. 138]. IBM Stuttgart.
- Mani, I., & Pustejovsky, J. (2012). *Interpreting Motion. Grounded Representations for Spatial Language*. Oxford University Press.
- Müller, S. (2013). *Grammatiktheorie* (2nd ed.). Stauffenburg.
- Petersen, W., & Petersen, W. (2015). Representation of concepts as frames. In T. Gamerschlag, D. Gerland, R. Osswald, W. Petersen, J. Skilters, F. Toccafondi, & G. Stemberger (Eds.), *Complex Cognition and Qualitative Science. The Baltic International Yearbook of Cognition, Logic and Communication* (commented reprint, pp. 43–67). Düsseldorf University Press.
- Pourcel, S. (2005). What makes path of motion salient? *Berkeley Linguistic Society (BLS)*, 30, 505–516.
- Pustejovsky, J., & Moszkowicz, J. L. (2011). The qualitative spatial dynamics of motion in language. *Spatial Cognition*, 11, 1–32.
- Randell, D. A., Cui, Z., & Cohn, A. G. (1992). A spatial logic based on regions and connection. *Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning*, 165–176.

- Slobin, D. I., Ibarretxe-Antunano, I., Kopecka, A., & Majid, A. (2014). Manners of human gait: a crosslinguistic event-naming study. *Cognitive Linguistics* 25(4), 701–741.
- Snell-Hornby, M. (1983). *Verb-descriptivity in German and English. A constrastive study in semantic fields*. Carl Winter Universitätsverlag.
- Takahashi, K. (1997). Verbs for global locomotory body motion in Thai. *Proceedings of the 16th International Congress of Linguistics (ICL '97)*. [<http://labo.kuis.ac.jp/private/kiyoko/GLBM.pdf>]
- Talmy, L. (1983). How language structures space. In H. L. Pick Jr., & L. P. Acredolo (Eds.), *Spatial Orientation. Theory, Research, and Application* (pp. 225–282). Plenum Press.
- von Stutterheim, C., Gerwien, J., Bouhaous, A., Carroll, M., & Lambert, M. (2020). What makes up a reportable event in a language? Motion events as an important test domain in linguistic typology. *Linguistics* 58(6), 1659–1700. <https://doi.org/10.1515/ling-2020-0212>
- Zlatev, J., Blomberg, J., Devylder, S., Naidu, V., & van de Weijer, J. (2021). Motion event descriptions in Swedish, French, Thai and Telugu: A study in post-Talmian motion event typology. *Acta Linguistica Hafniensia*. <https://doi.org/10.1080/03740463.2020.1865692>

## Web Resources

- Das Deutsche Referenzkorpus – DeReKo. IDS Leipzig-Institut für Deutsche Sprache Mannheim. [www.ids-mannheim.de/digspra/kl/projekte/korpora](http://www.ids-mannheim.de/digspra/kl/projekte/korpora)
- Duden. Cornelsen Verlag Berlin. [www.duden.de](http://www.duden.de)
- Wikipedia. Die freie Enzyklopädie. <https://de.wikipedia.org/wiki/>
- Wortschatz Leipzig. Universität Leipzig, Sächsische Akademie der Wissenschaften zu Leipzig, Institut für Angewandte Informatik Leipzig. [www.wortschatz.uni-leipzig.de/de](http://www.wortschatz.uni-leipzig.de/de)



# What Are Events?

---

**Abstract** This study reviews research on the conceptual structures and cognitive mechanisms that underlie the human ability to cope with the dynamicity of the world, a phenomenon often subsumed under the term event cognition. It identifies four distinct theoretical perspectives in the literature: the participant-based view, the boundary-based view, the object-states view, and the event-layer view. The basic ideas that constitute each perspective are outlined, and the methods and empirical data that are typically presented in support of each perspective are discussed. After an evaluation summarizing differences and similarities, the conclusion is that likely only a unification of the four approaches makes it possible to capture all aspects of the phenomenon, including those related to perception, conceptual representation and linguistic encoding, and thus be of value to researchers across different disciplines. The final section offers some ideas on what such a unified approach might look like.

**Keywords** Event cognition; thematic roles; event boundaries; object states; event layers; cross-linguistic differences

## Introduction

When people communicate with one another using language, they assume that their communication partners will understand more or less the same things by the words that are being said. In everyday life, slight variations in word meanings often go unnoticed. But in scientific communication, variation in the use of a specific term may have immense consequences. In the worst case, misunderstanding, misconception, and theoretical incoherence may emerge, even more so if the same term is used across multiple disciplines. However, investigating how a specific term is used by different researchers, acknowledging and incorporating terminological variation can have a more positive outcome, as well: a more fine-grained and inclusive understanding of a phenomenon.

One term worth reflecting upon how it is used across different disciplines is “event”. It frequently appears in the works of researchers investigating how

humans are capable of interpreting the ever-changing environment that they live in. Thus, it can be found in theories from a variety of fields that form (part of) the cognitive sciences: events play a role in research on perception, language, and memory. Although events may be treated as entities that happen among living organisms and inorganic materials in the world that surrounds us (Casati & Varzi, 2020), in research on cognition, events are taken as phenomena of the inner world. One of the issues in the cognitive sciences is how the mind generates and handles abstract representations of the outer world, whereas the ontological status of events is not necessarily the primary concern. The exact relationship between what is ‘within’ and what is ‘out there’ is apparently so complicated that thousands of years of philosophical thinking have not yet come to a definite, all-agreed-upon answer (Boden, 2006; Fuchs, 2017). For the purpose of this chapter, we will assume that an external stimulus can activate an internal representation and that an internal representation is critical for how the outer world is perceived and experienced at a given moment. So far, so good. But are event representations also evoked by internal simulation when one is contemplating experiences in the past, possible experiences in the future – or even imagining possible worlds while reading fiction? This paper aims to provide some cornerstones for the study of events in a way that reduces misunderstanding and thus provides a fertile ground for future research across disciplines.

If events are mental representations, how do they become available to real-time cognitive processing, what do they specify exactly, and how are they thought to play out in human cognition? These are the questions that we will address. Different theoretical perspectives can be identified in the literature, and each provides its own answers. Thus, we will start by introducing those views that we think can be grouped together and those that are unique enough to stand on their own. We will focus on the following four perspectives: the participant-based view, the boundary-based view, the object-states view, and the event-layer view. In the second part, we will summarize and evaluate the different views. We will try to point out some of the ideas that all approaches share, but also how they differ. In the last section, we will sketch in broad strokes how some of the differences may be overcome within a unifying framework. A disclaimer is in order, though: Due to space limitations, we can only present the core ideas of each of the four approaches, with only a limited level of detail. We hope that some of the references will help the interested reader to fill in the gaps that may arise.

Although it may be interesting for historical reasons, we do not think that it would be much of an advantage to tightly link every theoretical perspective we discuss to the specific discipline that it originated from because this might lead to the impression that the respective approach would be valid for that discipline only. Our goal here is to describe a broad spectrum of ideas

revolving around the theoretical concept of event representations, and since all approaches, at their core, try to capture how the mind represents and processes the dynamic aspects of the world, they all reflect different aspects of the “same cognition” – at least this is the perspective that we chose for this chapter.

## Four ways to approach event representations

### The participant-based view

Participant-based views typically attribute their main idea to Charles Fillmore’s (1968) *The case for case*. The theoretical core is evident in the following quote.

The case notions comprise a set of universal, presumably innate, concepts which identify certain types of judgments human beings are capable of making about the events that are going on around them, judgments about such matters as who did it, who it happened to, and what got changed. (pp. 45–46)

Note that what Fillmore here calls “case notions” is referred to as “thematic roles” in subsequent work by most authors. According to Fillmore’s conception, event representations comprise abstract atomic concepts. These concepts can be referred to by language and are thus represented separately from the domain of language. Each concept within an event representation corresponds to an entity that, in one way or another, participates in the represented event (hence the term “participant-based view”). Thus, two types of information are relevant in the participant-based view: the number of participating entities and how these entities are specified to relate to one another. For example, an event representation that may be referred to verbally by *The mother is waking up her girl* is a representation specifying two referents (*the mother, her girl*) as well as certain relations between them: one acts as an agent and a second one as a patient.

Dowty (1989) clarifies that Fillmore’s original conception has not been perceived consistently in subsequent research. There are at least two versions of the participant-based view as derived from Fillmore’s original theory. According to what Dowty labels the “thematic role type” conception, events are represented in terms of abstract role categories. If an event representation includes an agent, it refers to the same abstract agent concept as any other event representation that includes an agent. Similarly, if an event representation includes a patient, it refers to one and the same abstract patient concept as any other event representation that makes reference to a patient. Theories in this vein assume a finite set of abstract thematic roles from which every

event representation can be constructed. Specific configurations of thematic role types result in different event types. For example, if an agent and a patient are part of the event representation, this may constitute a ‘causative event’; if there is an agent, recipient, and a theme, the event will be of the ‘transfer event’ type; and if there is a moving entity, a path, and possibly a landmark or a goal, then we are dealing with a ‘motion event’. Since the assumption is that there is a finite set of thematic role types, there is a finite set of event types. Note, however, that no agreement has been reached on which roles exactly constitute the assumed finite set, or how any particular role type may be defined, so that it is specific enough to be distinguished from other role types and general enough to cover a broader range of cases (Dowty, 1991; Rissman & Majid, 2019).

Things are different in what Dowty (1989) calls the “individual thematic role” conception, according to which each event representation has its own configuration of thematic roles. In eating events, for example, one referent is assigned the ‘eater role’ and another referent is assigned the ‘what is eaten role’. In borrowing events, one referent is assigned the ‘borrower role’, one is assigned the ‘borrowee role’, and one is assigned the ‘what-is-borrowed’ role. According to this theoretical conception, a high degree of similarity in participant configuration is assumed to hold primarily for exemplars of the same event type, e.g., for all eating events or all borrowing events. At the same time, the possibility that abstract role types exist is not completely ruled out, but it is not the core assumption.

Even though the core idea of event representations being configurations of thematic roles originated in linguistics, it has been and still is very prominent in experimental psychology today, and the way in which respective empirical studies are designed reflect the differences between the two interpretations of Fillmore’s original idea.

Studies that are motivated by the “thematic role type” conception typically aim to answer questions regarding the psychological reality of role types. This is done by attempting to obtain empirical evidence for cognitive biases associated with the processing of abstract roles across different events (Rissman & Majid, 2019). One line of research, for example, employs eye tracking during verbal description tasks as well as non-verbal decision tasks using visual stimuli. These studies focus on how people visually process pictures showing objects that can be interpreted as event participants. Typical research questions in such studies are: Do people show starting point preferences when looking at the experimental stimuli? How fast can people detect a given role? What factors modulate participants’ performance (Dobel et al., 2007; Glanemann et al., 2016; Griffin & Bock, 2000; Hafri et al., 2013; F. Wilson et al., 2011). In general, the results from such studies show that people can very quickly detect or identify an object that carries a specific event role in

a depicted scene, that is, within a split second. This suggests that the interpretation of a visual scene as depicting an event is driven by the assignment of specific event roles to individual objects.

While thematic role assignment is apparently fast and automatic, not all roles seem to be processed equally: Some authors propose an agent-bias during processing, as agents have a special status in driving predictions about an event (Cohn & Paczynski, 2019; Kuperberg, 2021). Others propose that there is not only an agent bias but a cognitive hierarchy of event roles, placing agents before patients which are followed by goals and instruments. The hypothesis is that this hierarchy becomes apparent not only in frequency of mentions in free description tasks in which speakers can choose what to mention but also in preferences for the order of fixation, the time people need to move their eyes to the respective participant if the task demands it (F. Wilson et al., 2011) as well as in memory encoding and retrieval for different roles in the postulated hierarchy. So far, there is some limited evidence for such an event role hierarchy, although it is not clear whether this hierarchy is stable or situational. Recent studies suggest that it is malleable to a certain degree and dependent on various contextual factors such as priming (Saupe & Flecken, 2021).

Although experimental research has a strong bias towards English and a few other languages, the cross-linguistic studies available may provide a promising road to investigate how thematic roles cluster with respect to their linguistic expressions in different languages and to draw conclusions about underlying (universal) conceptual structures, i.e., event representations (Flecken et al., 2015; Gerwien & Flecken, 2016; Rissman & Majid, 2019; von Stutterheim et al., 2020). However, an important question still under debate is whether differences that may be linked to the analysis of linguistic structures across languages indeed reflect the way people represent events at a domain-general level, that is, independently from the verbal domain (Gennari et al., 2002; Gerwien & von Stutterheim, 2018; Ünal et al., 2021).

To investigate the relation between linguistic and non-linguistic event cognition, some studies focus on individuals who are not assumed to exhibit biases stemming from linguistic experience, like apes, infants or congenitally deaf persons who were not taught any conventional sign language (Goldin-Meadow, 2003). Findings from this line of research suggest that the existence of different participant roles in events is deep-rooted in cognition and that agent-based event cognition even predates human cognition (V. A. D. Wilson et al., 2022; Zuberbühler & Bickel, 2022).

Studies in line with the tradition of the “individual thematic roles” conception follow a research agenda that focuses on the specific content of an event representation rather than the abstract relations between event participants. If we assume individual thematic roles to refer to semantic features that can constrain or bias referent selection or highlight or induce certain properties of

the referent, then not every possible referent is a good candidate for a certain role, e.g., in an arresting event, a crook is less typically associated with the ‘arrestor-role’ than a police officer. The concept of a typical arrestor might even be automatically activated upon hearing the verb ‘to arrest’. Ferretti et al. (2001), for example, argue that events are represented as schemas that specify prototypical referent concepts. In a series of masked priming experiments, they showed that specific verbs can prime prototypical agent, patient, patient-feature and instrument concepts. Vice-versa, typical referents for different semantic roles, including locations, also prime typical verbs (McRae et al., 2005). Thus, event schemata in this perspective are not only abstract structural relations between participants, but also clusters of object and action features.

It is fair to say that participant-based views dominate experimental research on event representation and cognition today. Despite the theoretical controversies surrounding the definition of a finite set of event roles, experimental psychologists with different backgrounds seem to favor this approach. One reason might be that it is relatively easy to develop an experimental design. Another reason might be that the event roles conception offers a simple vocabulary to describe what parts constitute an event representation. And once a researcher has words for the parts of a whole, it becomes possible to study the processes that create the whole from its parts.

The participant-based view has gained influence in the computer science community as well, where it serves as the theoretical backbone to implement mechanisms that enable machines to achieve natural language processing tasks. To some extent in the spirit of the individual roles conception, projects like FrameNet (Fillmore & Baker, 2009; Johnson et al., 2016) work on establishing a database of abstract ‘frames’. These frames specify event types based on the relations that hold between referents that are interpreted to participate in the event.

### The boundary-based view

The boundary-based view departs from the perspective of a third-person observer, and it holds that events are established by unitizing the continuous perceptual flow of information into discrete event units. In this view, events are time intervals that are delimited by event boundaries. Whatever lies between two boundaries is represented as an event unit. The following quote from Zacks et al. (2007) illustrates the idea:

Thus, the system alternates between long periods of stability and brief periods of change. Periods of stability are perceived by observers as events and periods of change are perceived as the boundaries between events. (p. 275)

Event Segmentation Theory (“EST”, Zacks et al., 2007) is the most prominent representative arguing for the boundary-based view. Importantly, EST explicitly provides a mechanism by which humans create event representations; it thus explains how event representations emerge in real-time, that is, as resulting from automatic cognitive processing and associated neural activity. In EST, an event representation is called a ‘working model’, a term that highlights the fleeting, or transient nature of this specific type of representation. In short, a working model is a combination of abstract schema knowledge that is retrieved from long-term memory and the current input which the sensory system provides. A working model is kept activated in working memory, and once it is set up, a specific mechanism allows it to remain stable over some time. This is due to the assumption that new sensory input is predicted based on the working model and slight discrepancies between the model and the input are accommodated to some extent. However, if prediction error reaches a certain threshold, meaning if new input cannot be predicted well enough any longer, the current working model is abandoned and a new working model is established, again by retrieving a matching event schema from long-term memory. As studies suggest, this typically leads to the conscious perception of a breakpoint: one event has come to its conclusion and a new event begins (Zacks et al., 2001).

Experimental research grounded in the boundary-based perspective places great focus on the detection of event boundaries, namely by employing a task commonly referred to as ‘event segmentation task’ or ‘event unitization task’. Event segmentation is typically studied by presenting people with videos and asking them to press a button whenever one ‘situation’ ends and a new one begins (Newtson, 1973; Zacks, 2020). An analysis of when participants press the button while watching the videos reveals a relatively high agreement in where breakpoints are reported, although, of course, some variance between subjects arises. In same-group comparisons, this variance is attributed to the level of granularity that an individual chooses for detecting breakpoints. When choosing a fine-grained level for segmentation, subjects may detect many breakpoints, while choosing a coarse-grained level, fewer boundaries are reported in the same stimulus. A fine-grained or a coarse-grained segmentation can be induced by providing specific instructions. Importantly, breakpoints detected when choosing the coarse level coincide with some of the breakpoints that are reported when choosing a fine-grained level, highlighting that smaller event units are part of larger units. Such findings illustrate an important notion in event cognition research, namely that events can be seen as hierarchically structured. Macro-events comprise micro-events. Or the other way around: micro-events can be combined to form a macro-event (Bohnenmeyer et al., 2007; von Stutterheim et al., 2020; Zacks & Tversky, 2001).

When correlating data obtained in the behavioral button-press task with neuropsychological data, as obtained with fMRI and EEG, two important observations can be made: 1) there are brain activity correlates of event segmentation; and 2) the same areas are activated both in active event segmentation and in passive viewing, where subjects simply watch a movie without being instructed to perform an event segmentation task (Zacks et al., 2001). These findings suggest that event segmentation happens automatically during perception.

A large body of research has explored how event boundary processing relates to information processing as well as memory encoding and retrieval in different groups of participants (for a recent overview, see Zacks, 2020). In these studies, participants typically watch video clips and their recognition and memory for what was shown in the videos is tested under different conditions. As event boundaries trigger the updating of a working model, information that occurs at a boundary receives special processing. For example, people notice changes in actors' clothing better at event boundaries than in intervals between boundaries (Baker & Levin, 2015). Similarly, objects that occur at boundaries are recognized better than objects that occur between two boundaries (Swallow et al., 2009). Integration or binding of information among elements occurs primarily within a segment between two boundaries, so cueing one within-segment element benefits the retrieval of other pieces of information whereas an intervening boundary may prevent these benefits (Ezzyat & Davachi, 2011). Across different segments, memory favors information that helps to establish causal links (Radvansky, 2012). In general, segmenting helps people to better recall fine-grained information about a series of events (Gold et al., 2017) and reduces temporal compression in memory (Jeunehomme & D'Argembeau, 2020). Event segmentation has also been studied in young vs. older participants and in healthy vs. Alzheimer's patients (Zacks et al., 2006).

To what extent language plays a role in event segmentation is not well understood to date. There are several ways in which the structure and the lexical repertoire as well as conventions of language use (pragmatics) may show effects. On the one hand, the way in which event schemas form over time during cognitive development could be influenced by the language one speaks, because how the members of a given language community use their language to talk about events may impact what information is most frequently clustered and what information is kept separate while talking about events. This could be labeled a long-term effect of language. On the other hand, a short-term effect of language could be attested if people formed event units differently when comparing how they segment input while talking about it in contrast to when they do not (Gerwien & von Stutterheim, 2022). That language has an impact on unit formation is suggested by the findings of a study



by Gerwien & von Stutterheim (2018). They compared speakers of French and German in a verbal and non-verbal segmentation task and found that speakers of French were more likely to indicate event boundaries by button pressing when watching video clips showing a moving entity that changed orientation or direction in the course of the clips. Critically, French speakers also produced more assertions when describing the events spontaneously. The rationale of the study was based on the observation that speakers of so-called “verb-framed” languages, like French, must use individual verbs to refer to multiple path segments that a figure in motion traverses, while speakers of a “satellite-framed” language, like German, can combine multiple path segments in one assertion. However, the possibility that participants in the non-verbal task used inner speech to comply with the task could not be completely excluded, meaning that structuring the information for providing a button-press response could have been guided by internally creating linguistic structures. If that were the case, however, then the same argument could be made for any other event segmentation study.

To conclude our outline of the boundary-based approach, it should be acknowledged that compared to the participant-based view, which primarily focuses on the content of an event representation, the boundary-based view is much more concerned with the mechanism that creates event representations. Content-wise events are considered as having an internal structure, insofar as goals of agents and the outcome of actions play an important role (Kuperberg, 2021).

### The object-states view

According to the object-states view, events are not taken as stretches of time per se, as in the boundary-based view, but rather as stretches of multiple times – or rather: ‘time intervals’ – for the individual object concepts that comprise an event. Thus, like in the participant-based view, events are composed of sub-components. There are two theories that may be subsumed under the term “object-states view”, the Argument-Time Structure (“ATS”) theory proposed by Klein (Klein, 1999, 2010)<sup>1</sup> and the Intersecting Object Histories (“IOH”) theory, first fully spelled out in Altmann and Ekves (2019), but formulated in part already in Hindy et al. (2012). According to both theories, event representations specify how properties of objects develop over time. In the simplest case, an event representation specifies properties that are associated with only one object during an initial or ‘source’ time interval and during a resultant or ‘target’ time interval. The major difference between

---

1 Klein’s Argument-Time Structure conception was not intended to be a theory of event representation, but a theory capturing verb meanings and grammatical operators.

ATS and IOH is that IOH assumes that all time intervals between the initial and resultant state are part of the event representation (hence, “object histories”), whereas ATS does not make that assumption. Of course, event representations can be rich and specify states for more than one object. To illustrate the main idea from the perspective of ATS: In any event that represents some sort of transfer of possession, several objects co-occur in space and time. For example, in an event that may be referred to as *Mary threw John the ball*, there would be object concepts for *Mary*, *John*, and *the ball*. In addition, this particular event representation would contain an initial state for *Mary* which specifies the possession of *the ball* and a subsequent target stage at which the goal of the transfer – *John* – has come into possession of *the ball*. The specification of the properties that define each object state and how these states are associated via temporal and causal relations constitutes the minimal content of an event.

Since the argument-time structure of an event is a reductionist analysis of the descriptive and structural content, it is not a part of the ATS framework to account for transitional gaps between initial and final object states (e.g., a trajectory of a ball traveling through space and time before reaching a goal). IOH, on the other hand, posits a fine-grained view on object state representations by including all time intervals between the initial and resultant stage, thereby including transitional states. Both theories recognize that events are ensembles of objects (understood as entities with clear spatio-temporal boundaries, including living organisms) that undergo some form of state change over time. However, IOH explicitly states that our understanding of events implies the simultaneous activation of all object states associated with a specific event, i.e., the object histories in their entirety. In perception, this means that all object token states from the perceptual input must be bound into a coherent object representation and mapped onto semantic memory. From this view, generalizations such as participant roles are not the primitive components of event representation. Participant-roles result from patterns derived on the basis of perceived object changes.

The assumption of multiple state activation is backed up by both a priori arguments as well as experimental evidence. From a theoretical point of view, the activation of an object *history*, as opposed to the activation of a single perceptually salient or linguistically highlighted object *state*, is necessary in order to understand that a change has occurred. For instance, the content of the sentence *The burglar opened the door* is only recognized as an event as long as a previous state of *the door* (*‘not open’*) is part of the representation. The reasoning behind multiple state activation can be found in the framework of ATS as well. Here, object state activation is described as logical dependencies (so-called ‘H-connections’, named after David Hume), which imply that one object state would not be able to constitute a target state unless a previous

initial state is counterfactually implied. A simple object state described by a sentence such as *The door is open*, in which *the door* has a single attribute (*'open'*), is not recognized as an event since it does not imply a counterfactual relation to a previous state.

There is experimental evidence to support the idea of multiple object state activation. This concerns the following prediction: since events are understood as ensembles of object histories – meaning that representing an object implies representing multiple states of the same object at the same time – cortical activation patterns representing different object states must compete when a contextually appropriate object state is retrieved. In a study using fMRI, Hindy et al. (2012) found activation patterns suggesting such competition effects of different object states being simultaneously activated. In this study, participants read sentence pairs describing two different events containing the same physical object with the task to decide if the two sentences formed a coherent mini discourse or not. The first sentence described an affected object either as minimally changed (*The squirrel will sniff the acorn*) or as substantially changed (*The squirrel will crack the acorn*). The second sentence either described a preceding or proceeding interaction with the object (*But first, it will lick the acorn* or *And then, it will lick the acorn*). In order to understand the second sentence, a decision between retrieving the initial or target state of the critical object (acorn) had to be made. In this example, the acorn must be retrieved either in a sniffed or un-sniffed (minimal change) or in a cracked or uncracked state (substantial change). Whereas sniffing an acorn would leave the substance of the acorn intact, cracking an acorn would change it. Since the comprehension of an interaction between the squirrel and the acorn would imply deciding which object state of the acorn is contextually relevant, a subsequent interaction involving a substantial change in state of the acorn would induce a greater semantic conflict than a minimal change. This means that, given our world knowledge of the affordances of an acorn, the semantic conflict would then arise in the situation in which the squirrel would lick a cracked acorn. The results of the study showed activation patterns in the substantial change-condition indicating competition in the retrieval of different object states. Further experiments confirmed that the effect is not due to the processing of specific lexical elements, but indeed to the change of a specific object in the context of the event (Experiment 2 in Hindy et al., 2012; Solomon et al., 2015). In addition, results from two further studies suggest that the competition effect is indeed linked to the subsequent reference to the object (Kang et al., 2020a; Prystauka, 2018). Reaction time studies by Kang and colleagues (2020b) and Horchak and Garrido (2021) using a picture-sentence matching task confirmed that after a sentence implying a change in object state, initial and resulting state remain activated regardless of the object being mentioned again in a second sentence. The underlying

neural mechanisms were investigated in more detail by Hindy et al. (2015) in an fMRI study. They concluded that different object states are represented as being different in the primary visual cortex, but as the same stable object in the left ventral posterior parietal cortex, and that the mechanism observed in the study by Hindy et al. (2012) supports the selection of the sensorimotor representation relevant in the current context.

Another line of research is less concerned with the degree of change that an object undergoes, but rather with the complexity that arises per se if a verb denotes multiple object states. Gennari & Poeppel (2003) compared eventive (multiple object states) and stative (single object state) verbs using reading times and lexical decision reaction times and found that eventive verbs take longer to process, which they explained by the more complex event structure. Gerwien (2011) followed a similar logic and compared reading time for intransitive verbs that do or do not denote a change of state. Results show longer processing times for more complex verb meanings.

In summary, the object-states view defines events with focus on their unfolding over time. Unlike the boundary-based view which is concerned with the cognitive mechanisms segmenting events into units, the object-states view sets out to explain how the representational content of these units becomes available over time. Representational content can be described as bundles of object states co-occurring in spatio-temporal proximity. This view on events can be described by principles laid out by ATS and IOH. While ATS models the minimal descriptive event content (i.e., object states) that can be referred to by verbs as well as the logical and temporal relations between object states, IOH focuses on perceptive mechanisms of binding multiple object states into coherent representations. The latter theory then shifts focus from generalized static knowledge onto the neurocognitive processes involved in abstracting generalized representations from primitive building blocks during event conceptualization.

### The event-layer view

The event-layer view starts with the observation that things in the world frequently happen in parallel. Bennett (2002) refers to this observation as events being ‘thick’. To illustrate this view, one can imagine that when a leaf falls from a tree, the leaf would be moving downwards while possibly also rotating around itself. An object can be in motion on some path, for example, moving downwards, while at the same time exhibiting a specific manner of moving, for example, rotating. Importantly, the relation between falling and rotating in this example is not hierarchical in nature. Falling is not a subevent of rotating and rotating is not a subevent of falling. While events can be and often are hierarchically embedded in a taxonomy of causal links (see Goldman,

1970; Löbner, 2021 and “boundary-based view” section), the event-layer view captures a different kind of complexity in event representation, i.e., parallel occurrence with no explicit causal relation.<sup>2</sup> The relevance of parallel occurrence for questions of event representation is, for example, laid out in the work of von Stutterheim and colleagues (Gerwien & von Stutterheim, 2022; Gerwien & von Stutterheim, 2022; Lambert et al., 2022; von Stutterheim et al., 2020; von Stutterheim & Gerwien, 2023). The main idea behind what this research group studies under the term ‘event layers’, is that an observer of a scene selects one specific dimension – one ‘event layer’ – as a starting point for the construal of an event representation. The choice in turn has consequences for unit formation. Imagine a person running toward a train station and then entering it. Choosing the ‘path layer’ for event construal would lead to two event units in this case, one that represents the phase of approaching and one that represents the phase of entering the train station. Choosing the manner layer, on the other hand, would yield only one event unit which would represent the phase during which the figure continuously exhibits the same manner of motion without changes (e.g., running). Von Stutterheim and colleagues explore the event-layer view in several studies, in which they collect spontaneous verbal descriptions from speakers of different languages who respond to short video clips of real-world situations. The analyses of the verbal responses reveal patterns of what information speakers select for expression, what information they omit, and over how many clause-sized linguistic units (assertions) they distribute the information they provide. The authors interpret these patterns in verbal descriptions as to reflect preverbal event representation, that is, the representation of information that feeds into linguistic encoding.

Note that the event-layer view may not be restricted to the domain of motion events. Viewing the construal of an event representation as selecting one privileged layer as a starting point may apply to causative and other event types as well. Take a sentence like *Someone is folding a paper airplane*. Using this sentence in English refers to a manner in which an actor handles a piece of paper (folding), and at the same time it refers to the changing of a sheet of paper from one shape into another, which can be described as the intentional goal of the actor. While there is nothing special about expressing information

---

2 The existence of non-causal relations between simultaneously occurring actions has been identified by Goldman (1970) (“simple generation”; p. 31), exemplified by the parallelism of the acts of *hitting the tallest man in the room* and *hitting the wealthiest man in the room*, by which the same person is being referred to. In line with Goldman, Löbner (2021) explains such relations as a “constellation of facts” (p. 270). However, there is no further elaboration on how non-causal relations between parallel events occupying the same spatiotemporal zone should be modeled in abstraction from human action.

on the intentions of the actor and the manner by which the intentional goal is being reached for speakers of English or German, many languages do not allow speakers to refer to both layers – the manner layer, and the intentional layer – at the same time, i.e., in a single assertion. For example, to express the same information conveyed by the example above, a speaker of French would have to use the verb *plier* (to fold) and the verb *faire* (to make). If she intends to produce only one assertion, either the folding (manner) or the paper plane-making (intentional goal) needs to be omitted.

The event-layer view, as far as it has been made explicit in the published work of von Stutterheim and colleagues up to this point, should not be regarded primarily as an attempt to identify a finite set of layers that may comprise ‘thick’ events. So far, the types of layers that may be identified include the path (in motion events), the manner, the intentional, and a ‘witness’ layer (von Stutterheim et al., 2020), although the authors emphasize that this list is not exhaustive. Also, linking specific information selection patterns that may be found in different languages with specific features of the grammar and lexis in those languages is in our view not the most relevant aspect of the approach. What we think is worth underlining here is that events are regarded as representations that are *construed* by actively directing attention to information, either by the need to fulfill the current task requirements or by a default mechanism if there is no specific task. Thereby, some information receives a privileged status, whereas other information is defocused. Note that only after the choice for a specific layer or layers has been made, it can be determined which objects will be represented as participating in a given event and which objects will not. If we use the terminology of the participant-based view, selection of an event layer determines whether an object may be the agent or a moving entity, or whether an object will be a theme or an instrument, and so on. The choice for the layer determines what relations are relevant between which objects. Similarly, from the perspective of the object-states view, choosing one layer determines which object and which object states will ultimately be part of the event representation. Thus, assuming event layers, at least as an epistemological tool in the study of event representation, makes it possible to pre-structure the seemingly infinite possibilities of what information will be represented temporarily and what information will not be processed further.

Under a set of rather specific assumptions about the general architecture of the cognitive system and corresponding assumptions about the levels at which information are represented, i.e., a distinction between verbal and non-verbal levels of representation (Lupyan, 2012; Paivio, 2014; Van Dijk & Kintsch, 1983; Wolff & Holmes, 2011), the event-layer view may seem to imply that there is some additional, maybe even a ‘richer’, representation that provides the ‘material’ from which information is selected and represented specifically for

verbal expression. If one makes such assumptions, the question may seem justified as to whether speakers with different languages do indeed represent different event units at a non-verbal level of representation, or whether they simply represent the same events, but refer to them in different ways. We will address the question of how many levels of representation researchers assume in some more detail below. As far as the event-layer view proposed by von Stutterheim and colleagues goes, there is only one level of representation, and that representation holds information as provided by the sensory systems and information from activated event schemas, which to some extent are modulated by language (see Gerwien & von Stutterheim 2018, and above). Support for this view comes from studies on bilinguals and highly advanced L2 speakers, which suggest that some of the patterns in information selection during event construal prevail when participants use a language other than their native language to verbally respond to visual stimuli. Thus, what drives ad hoc event construal for verbalization is not necessarily influenced by the grammar and vocabulary of the language currently in use but may rather be driven by deeply-entrenched patterns of attention allocation that guide the information selection process and that have formed as a consequence of native language use. Another piece of evidence comes from the already mentioned study by Gerwien & von Stutterheim (2018), where the main manipulation concerned whether a person or inanimate object did or did not change direction / orientation in the course of short video clips. Results showed that French speakers were more likely to produce multiple assertions than German speakers in a verbalization task, and that they were also more likely to indicate an event boundary by button-pressing in a non-verbal event segmentation task. Von Stutterheim and colleagues argue that speakers of different languages make use of what they call ‘attentional templates’ that guide information selection and consequently the verbal encoding process. In their view, these attentional templates are stored as event schemata in long-term memory. Which of the available templates is used as a default in a given situation may vary from language to language (Gerwien & von Stutterheim, 2022; Lambert et al., 2022).

## Summary and evaluation

In the last sections, we provided an overview of four perspectives on event representation. In the participant-based view, events are defined by the number of referents participating in the represented event and the relations between them. In the boundary-based view, events are discrete units segmented out of the continuous perceptual stream. In the object-states view, events are seen as different states of objects associated with one another. And in the event-layer view, events comprise different qualitative dimensions

linked to one or several entities. Given these apparently quite distinct ideas on event representation, the question arises: Do all approaches mean the same when using the term “event”?

All four approaches assume that, at a theoretical level, events have ‘a right to exist’ as conceptual representations, next to other representations such as object representations. In this, all four approaches take events to be cognitive units. Considering events as cognitive units implies that, at least at some point during real-time processing, they are available to the cognitive system as a whole. Similar to objects, which may be analyzed as consisting of subparts – a human body for example comprises a torso, a head, legs and arms –, but which, for the sake of being interpreted as a whole, appear to be integrated at a given moment in time, events also integrate components to form a whole. This has implications for questions related to the processing, memory, and verbal encoding of events, as well as to the acquisition of event processing abilities during cognitive development and the potential loss of these abilities in cognitive decline.

The four approaches, however, differ with respect to the ‘size’ of an event unit, that is, regarding how many components, and thus, how much information can be integrated to form one unit. On the one hand, the participant-based view, the event-layer view and Klein’s Argument-Time-Structure (ATS) theory, as one representative of the object-states view, tend to restrict the amount of information that can be integrated to form an event to what can be expressed in one sentence. That is, while the duration of the temporal intervals that constitute events can vary freely – potentially ranging from something that can happen in the external world in a split second to something that can take place over minutes, days, months or even millenia –, the subparts that are to be specified and integrated depend on, and are restricted in number by what can be expressed in a sentence that includes a specific verb. The main components are the referents that are directly associated with the verb as the verb’s syntactic arguments. On the other hand, neither the boundary-based view nor Altmann and Ekves’ Intersecting Object Histories (IOH) theory in principle assumes any specific unit size in terms of the number of subcomponents that may be integrated to form one event unit. In IOH, units are defined by binding mechanisms, including the binding of objects to event schemata. Thus, whatever components the binding mechanisms are applied to will be part of the event unit. In the boundary-based view, units are defined by perceptual boundaries at one of potentially many different (hierarchical) levels whereas there are no principled restrictions as to the number or type of subparts.

Importantly, each perspective acknowledges the need to be able to explain why people can recognize events as being in a certain way similar to events that they have encountered previously and why people use similar



linguistic devices to talk about similar events. Thus, all approaches make a distinction between abstract event knowledge on the one hand and specific or instantiated event representations on the other hand. Obviously, it depends on the theoretical approach of what can be described as the content of an event representation to make explicit what ‘abstract event knowledge’ consists of exactly. Abstract knowledge may be taken as knowledge about abstract event roles or as knowledge about prototypical participants of types of events, including not only typical animate objects (agents and patients), but also typical locations and typical additional objects, such as instruments. Or abstract knowledge may be knowledge about how a person typically reaches a specific goal, e.g., knowledge about how an agent’s intentions play out. Or abstract event knowledge may be viewed as knowledge about how features of objects typically change over time.

Given that all approaches make a distinction between abstract event knowledge and specific instantiations of that knowledge, all four approaches, implicitly or explicitly, locate event representations in working memory. Note that we mean here the “generic definition” of working memory as provided in Nelson Cowan’s widely perceived review article “The many faces of working memory and short-term storage”, according to which the term refers “... to the ensemble of components of the mind that hold a limited amount of information temporarily in a heightened state of availability for use in ongoing information processing.” (Cowan, 2017). If it is acknowledged that events are units in working memory, then events are fleeting, or transient representations. Events are representations at a specific moment in time that vanish, unless kept activated by some cognitive mechanisms. However, the literature is not too explicit regarding whether the term ‘event’ should be restricted to this notion, or whether the term should be applied also to representations stored in episodic memory, i.e., to knowledge of specific personal experiences that have been represented as specific events at some point in the past, but that is not currently activated. If event representation is considered as a constructionist process, there seems to be no need for that, because recalling a personal experience could simply be thought of as (re-)creating a specific transient event representation from smaller pieces of information.

Although to different degrees and in different ways, the four approaches imply that the formation of an event representation in working memory requires a control mechanism by which relevant information is selected and shielded from interference by irrelevant information. Despite the fact that the term ‘attention’ is heavily debated in current research, with some authors even calling for its abandonment (Anderson, 2011; 2023), we will nevertheless use it here. In this, we view working memory as a storage component and attention as the processing that acts upon the temporarily stored information. It is hard to imagine how the construal of a cognitive unit from ‘smaller

parts' by means of integration, and how keeping the resulting representation activated over some time could be achieved without assuming any type of an attentional control mechanism. To consider attention a necessary cognitive prerequisite to study event representation and event cognition is imperative to tackling a variety of issues that could not be resolved otherwise. First, attention prevents random shifts between different granularity levels during event unit formation (see section 'The boundary-based view'). Without attending to one particular level of granularity at a given moment in time, it would be impossible to interpret a dynamic scene coherently or to find words to talk about it. This is also evident from the fact that humans do not seem to automatically establish every relationship between a given event and its embedding in, or relation to, the global affairs of the world at all times (cf. the micro-/macro-event distinction). Second, attention is required to select relevant information in the sense that the layer-view suggests. For example, while observing traffic with the goal of crossing a street, one most likely will attend to the movement of vehicles (the path layer) and not that the drivers of the vehicles have intentions of moving their vehicles to certain destinations (the intentional layer). Similarly, the manner in which a particular action is performed may not be represented obligatorily as part of an event: If you hear someone say that person A woke up person B, you may represent the change of state associated with person B but not how that change of state was brought about, e.g., by hitting a metal pot with a wooden spoon, or by poking, whistling, or calling a name. Choosing a hierarchical or qualitative level is relevant for event construal both based on the immediate visual input as well as for event construal from memory. Thus, even though at all times multiple dimensions are available to extract information from in order to form an event unit, the selection depends on the goal or task at hand, for which on theoretical grounds the assumption of an attentional selection mechanism seems to be required. Yet another reason for why an attentional control mechanism must be assumed to make event representation possible can be subsumed under the term 'perspectivization'. If working memory, as a temporary storage contains the components that make up the event, then these components can be 'profiled' in different ways. In a specific event, the focus may be on the causer or on what is being caused, as illustrated by comparing the event representations referred to by *Tom was woken up by a loud noise* and *The loud noise woke up Tom*. A similar case is illustrated by comparing what is being referred to by *The hunter is chasing the deer* and *The deer is fleeing from the hunter*. Since information selection, information integration and information manipulation are inherent to some extent in all of the presented approaches, some theoretical assumptions about temporary information availability (storage) and processing (attention) must be made.

## Levels of representation

How many levels of representation must be assumed to understand event cognition as the fundamental human ability to structure information in order to make sense of it and to communicate about it, no matter whether the origin of the to-be-structured information is perception, episodic memory, imagination, or language? It must be attested that there is no consensus on whether the modality independence implied by the question is a valid assumption at all. There are at least two reasons: One, researchers have been approaching event representation from individual research fields, and therefore from different directions, e.g., from visual perception or from language. Two, some researchers study events in isolation and others in context. The first reason leads to two very different views: On the one hand, there is the assumption that event representations are modality-specific, that is, event representations construed for linguistic encoding and events resulting from (visual) perception are not identical (e.g., Papafragou, 2015; Papafragou et al., 2008). On the other hand, there is the assumption that event representations are not specific to one modality, that is, the view that events are represented at a domain-general level. The second reason concerns the relation between events and another representational device termed ‘mental models’. We will start with addressing the latter.

While many researchers distinguish between events on the one hand and mental models (Johnson-Laird, 1983) or situation models (Van Dijk & Kintsch, 1983) on the other hand, a few researchers in the domain of event cognition assume a homology between situation/mental models and events. For example, Speer et al. (2007), in a study based on EST, write: “As events structure visual activity, situation models necessarily structure narrated activity”. First, it is important to note that “. . . although there were many differences in the ways we [Van Dijk/Kintsch and Johnson-Laird] used the notion of a model [in the terms ‘situation model’ and ‘mental model’], the fundamental idea was the same.” (Van Dijk, 1995). Thus, what is said in the following regarding the more general theoretical concept of a ‘mental model’ applies to the somewhat more specific concept of a ‘situation model’, as well. A ‘mental model’ is traditionally seen as a representation that is a complex mental simulation of relevant aspects of the world at a given point in time, which integrates different pieces of information from multiple sources, including information provided by the perceptual apparatus and information from long-term memory (scripts and schemas). This means that mental models are modality-independent in the sense that they combine information from different modalities into one representation. Note that it is a separate question whether this implies a multimodal (modality is preserved) or amodal (modality is not preserved) format. Humans use mental models to tackle all

sorts of higher-order cognitive tasks, such as text/discourse comprehension (Kintsch & van Dijk, 1978; Van Dijk, 1995; Van Dijk & Kintsch, 1983), reasoning and problem solving ( Craik, 1943; Johnson-Laird, 1983). Mental models are updated whenever new information becomes available, and thus, a mental model is typically conceptualized to integrate multiple individual events as they become available as new perceptual input or from memory retrieval. This implies that, in theory, event representations, as they are understood by the majority of researchers in the domain, and mental models cannot be representations at the same level. There cannot be a homology. But how, then, do they relate to one another? On the one hand, events can be understood as some of the building blocks of mental models, though, mental models integrate other ‘non-eventive’ information such as visual properties of a scene like the color of the sky, or factual knowledge such as knowing that it is dark at night as well. In other words, event representations could be seen to constitute some of the ‘material’ that mental models are made of. On the other hand, a mental model may provide the information from which one or several events can be constructed, e.g., for verbalization. If you have experienced something extraordinary yesterday, then what happened to you will be represented as a whole in the form of a mental model. If you want to tell your friends about it, you may select relevant pieces of information from your mental model and create one or multiple event representations that allow you to refer to different parts of it. Whether events are the building blocks of a mental model (among further non-eventive information), or whether a mental model provides the information for event construal (among further non-eventive information) simply depends on the perspective: Perceiving information that can be interpreted as an event can create or amend the current model, whereas, if the cognitive task is to evaluate aspects of the current mental model, e.g., in decision making, or to communicate about what is in the model, an event representation may be constructed from the information in the model. From this perspective, events are cognitive devices that interact with mental models in the service of the current task.

We now return to whether event representations are modality-specific or modality-independent. The boundary-based view (EST) and IOH – one of the representatives of the object state view – both focus on the mechanisms at play that underlie event representation. In this, both approaches do not explicitly differentiate between any modality-specific levels of representation. As Altmann and Ekves (IOH) put it: “There is little difference between directly experiencing [an event] and learning of it through language; yes, there are differences in detail (and goals), but by ‘little difference’ we mean in respect of the mechanism by which the tokenized representations come about” (Altmann & Ekves, 2019). In contrast, the large majority of researchers adopting the participant-based view assumes that events are represented at

multiple levels of processing and that it is possible to isolate different representations as the result of different processing steps, e.g., in the course of visually perceiving an event to the linguistic encoding of it (see section ‘the participant-based view’). To illustrate this way of thinking once more, imagine a situation in which someone is presented with a video clip with either the task to describe it (‘verbal task’), or to make some judgment about it or memorize it for later recall (‘non-verbal task’). In each of these scenarios, visual information first needs to be taken in via the sensory system. Some researchers assume that the result of information uptake leads to a representation of the visual input and based on that representation, further task-dependent representations may need to be generated. In the case of the verbalization task, the task-dependent representation would correspond to a semantic structure serving as a compatible input of the linguistic encoding system (following Levelt [1989], a so-called ‘message’). In the case of a non-verbal task, no such representation is assumed to be required. The comparison of eye movement patterns registered during such verbal and non-verbal tasks have been reported to show differences (Griffin & Bock, 2000; Papafragou et al., 2008). In addition, differences arise when eye movement patterns are compared in verbalization tasks between people with different native languages, whereas differences are absent when comparing groups of speakers with different languages in non-verbal tasks. Such findings are often interpreted as evidence for the ‘multiple-levels’-view (Papafragou et al., 2008; Trueswell & Papafragou, 2010). However, many have pointed out the methodological pitfalls associated with the goal to obtain evidence for non-verbal event representation in such a way, as well as problems with the underlying theoretical conception. For one, since the assumption is that different levels of event representation exist independently of different tasks, and different tasks only serve the purpose to tap into these different levels representations, it is unclear how one can dissociate between task-effects and ‘representation-effects’, so to say. It does not seem unreasonable to assume that the mind can generate representations specifically for the tasks at hand, which does not necessarily imply different levels of representations. An in-depth discussion of all issues is not possible here due to space-limitations, but we refer the interested reader to, for example, Wolff & Holmes (2011), Lupyan (2012) and Gerwien, von Stutterheim & Rummel (2022).

If one appreciates the theoretical concept of mental models and how events can be conceptualized to relate to them, namely in the form of transient representations that can operate on the information in the mental model, then the question whether events are represented at multiple levels of processing loses its theoretical relevance. Event representations are essentially associations of different pieces of information, which may be available in a multi-modal representational format. However, a different question may be

formulated: Is information that serves as input to the mental model already structured in the format of an event before its integration, or are events only formed within the mental model?

## What may a unified theory look like?

All four views that we chose to present here highlight different aspects of event representation as a cognitive phenomenon. Thus, all four approaches potentially specify the ingredients for a unified theory that can provide both a universally applicable format of event representations as well as the processing mechanisms involved in event cognition.

One starting point towards a unified approach would be to translate those ideas that primarily focus on the content and structure of event representation into a common format. One way to do so would be to assume that, in terms of their contents, events are nothing more but representations in which attributes of objects are attended to over a period of time. In this, the term ‘attribute’ applies to all kinds of inherent (color, shape, animacy, etc.) and extraneous (conceptual salience, intention to reach a goal, etc.) features that can be linked to objects. This includes features that can be subject to sensory perception (e.g., motion, change of integrity) and features that can be assigned (e.g., volition, intention). For the sake of the argument, let us refer to the first as ‘attribute perception’ and to the latter as ‘attribute assignment’. In both cases, attributes define a domain for attention allocation. By attending to an attribute of an object over time, the values of that attribute may either change or stay constant. One may refer to this as ‘attribute value tracking’. An event representation must at least bind two values of the same attribute to one object. Note that parts of this idea were laid out in Miller & Johnson-Laird’s influential book *Language and Perception* (1976). The general view just outlined may provide the basis for everything that can and must be said about the content of an event representation, even for very complex ones. In addition, it offers several advantages, which we will describe in more detail below.

Second, it would be necessary to adopt and adapt the mechanism laid out in the Event Segmentation Theory, that describes how generalized event knowledge is activated to create a ‘working model’, to the idea of ‘attribute tracking’. This should not be too complicated as quality changes (changes of attribute values) are already taken as what determines the deactivation of a current and the activation of a new event schema in EST, i.e., what constitutes the perception of a boundary between two events. However, previous research conducted from the perspective of EST has been more concerned with how the mechanism affects the construction of situation / mental models

rather than event representations in the sense we laid out here. This is where adaptation would be required.

Let us return to the idea of attribute assignment, attribute perception, and attribute tracking. First, allowing both attribute perception and attribute assignment as equal cognitive operations relevant in constructing event representations opens up the possibility of bottom-up and top-down driven modes of event cognition. In other words, it allows perceptual features to be the starting point to generate an event representation, and at the same time, it leaves room for conceptual guidance. This view allows interdisciplinary theories of event cognition to overcome the great (artificial) divide between perception and cognition. If attribute tracking is assumed to be at the heart of event cognition, it should not matter too much whether attention to attributes is *triggered* by perception or whether it is *assigned* driven by a task-goal. The resulting representation can have the same format.

Second, attending to object attributes over time (as opposed to viewing them as static features) is what theoretically differentiates object representations from event representations. Identifying the tracking of attribute values over time as the core of event cognition allows us to do away with the distinction between “events in the narrow sense”, where there needs to be a qualitative change (‘to wake up’), and “events in the broader sense” (‘to sit at the table’), where an actual (perceivable) change is not the defining criterion. Allocating attention to an attribute value during a time interval without identifying any change would allow to create a dynamic representation, meaning an event, and not a representation of a state of an object.

Since in many cases event representations integrate the tracking of attributes of *multiple* objects over time (e.g., ‘event participants’), an explanation is required of how this is achieved. One possibility is to assume pre-made structured representations to which attributes of multiple objects can be linked as elements of that structure – an idea which is in part laid out in the Argument-Time Structure theory (ATS, Klein, 1999). It may be possible to identify a finite set of time interval configurations to which objects and object attribute values can be bound, with the least complex one comprising only one interval for one object, and more complex ones comprising multiple intervals for multiple objects. It seems likely that the complexity of time interval configurations has an upper limit, which is imposed by the resource limitations of the cognitive system, most importantly the limit of items that can be maintained in working memory simultaneously. Assuming pre-made structured time interval configurations to which relevant object attribute values can be linked provides an easy solution to the problem of how the tracking of attributes of multiple objects results in an integrated representation. Figure 1 summarizes our proposal.

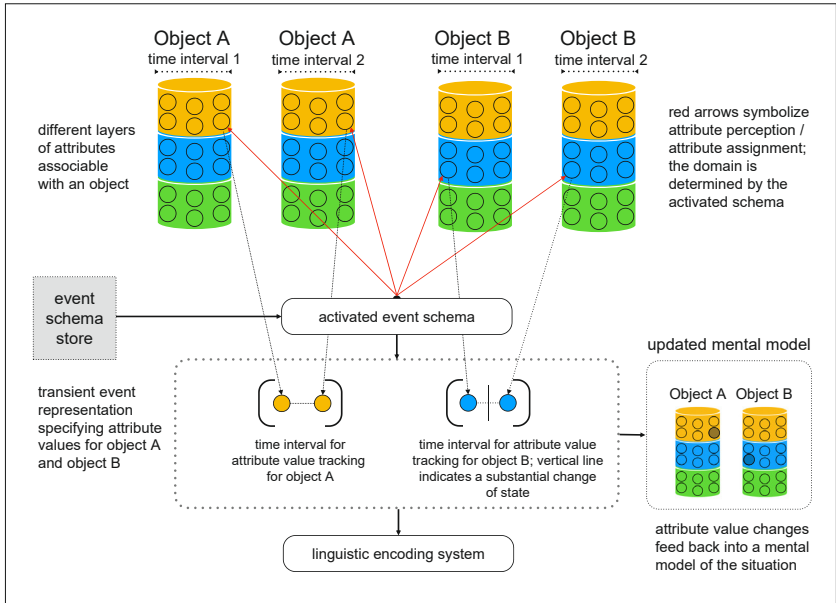


Figure 1. Illustration of a unified conception of event representation; the top part illustrates attribute perception, attribute assignment, and attribute value tracking, which all interact with event schema knowledge (red arrows); the area with the dotted contour illustrates an event representation; the right part illustrates how a specific event representation can update a mental model.

## Conclusion

We can't do without events, or can we? What researchers attempt to capture with the theoretical concept behind the term 'event' is fundamental to human cognition: the ability to survive and live in an ever-changing environment. The investigation of how humans perceive, remember, and communicate about events, and how these skills are acquired, however, poses some serious challenges to the interdisciplinary research of the phenomenon. As the mind (and its implementation in the brain) has been studied by different disciplines for almost two centuries now, each discipline brings its own epistemological instruments to the table – methodological and theoretical – which inevitably leads to a situation where the same phenomenon is latently associated with different discipline-specific concepts. For example, one cannot study events in linguistics without acknowledging basic linguistic notions such as the fact that words are combined into clauses and sentences, which provides the linguistic units onto which events must be mapped. Similarly, if events are studied in cognitive psychology, the processes by which the mind generates event representations must be linked to the assumed general



cognitive architecture, including core components such as the so-called “executive functions” (e.g., working memory / attention).

At some point during the preparation of this chapter and the discussions that went along with it, we suspected that the term ‘event’ suffers from what is known as ‘conceptual fragmentation’, as many other theoretical terms especially in cognitive science, like for instance the terms ‘representation’, ‘(working) memory’, ‘attention’, and even ‘cognition’ itself, among others. Conceptual fragmentation refers to cases where “... (i) a certain term, originally widely assumed to enjoy a single meaning, has been found to have multiple distinct meanings no one of which is privileged, and (ii) different definitions are adopted for different theoretical uses.” (Taylor & Vickers, 2017). However, if our suspicion was correct, we have hopes that different disciplines can find common ground regarding the concept of ‘events’. We think that it is possible to develop converging ideas on the underlying theoretical concept and use the term in a coherent way. Our goal here was to pave the way for a better understanding between disciplines by trying to extract four main approaches to event representation, crystallize each one’s core ideas, and so, for one, create awareness of the different facets of the phenomenon, and two, outline how it may be possible to combine them into a unified framework.

To conclude, event representation is a research topic that cuts through all aspects of cognition. Therefore, it is an arena where different disciplines can come together to learn from each other, present their unique perspectives, but also to solidify or challenge some of their specific notions and assumptions. Theoretical concepts that do not (fully) work in explaining event cognition may require re-evaluation. Finally, events are an interesting field for evaluating meta-questions in the field of cognitive science, such as representationalist and anti-representationalist views on cognition (and all positions in between).

## References

- Altmann, G. T. M., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review*, *126*(6), 817–840. <https://doi.org/10.1037/rev0000154>
- Anderson, B. (2011). There is no Such Thing as Attention. *Frontiers in Psychology*, *2*. <https://doi.org/10.3389/fpsyg.2011.00246>
- Anderson, B. (2023). Stop paying attention to “attention.” *WIREs Cognitive Science*, *14*(1). <https://doi.org/10.1002/wcs.1574>
- Baker, L. J., & Levin, D. T. (2015). The role of relational triggers in event perception. *Cognition*, *136*, 14–29. <https://doi.org/10.1016/j.cognition.2014.11.030>
- Bennett, J. (2002). What events are. In R. M. Gale (Ed.), *The Blackwell Guide to Metaphysics* (pp. 43–65). Blackwell.

- Boden, M. A. (2006). *Mind as machine: A history of cognitive science*. Oxford University Press.
- Bohnenmeyer, J., Enfield, N. J., Essegbey, J., Ibarretxe-Antuñano, I., Kita, S., Lüpke, F., & Ameika, F. K. (2007). Principles of event segmentation in language: The case of motion events. *Language*, 83(3), 495–532. <https://doi.org/10.1353/lan.2007.0116>
- Casati, R., & Varzi, A. (2020). Events. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2020/entries/events>.
- Cohn, N., & Paczynski, M. (2019). The Neurophysiology of Event Processing in Language and Visual Events. In R. Truswell (Ed.), *The Oxford Handbook of Event Structure* (pp. 623–637). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199685318.013.26>
- Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin and Review*, 24(4), 1158–1170. <https://doi.org/10.3758/S13423-016-1191-6>
- Craik, K. J. W. (1943). *The Nature of Explanation*. Cambridge University Press.
- Dobel, C., Gumnior, H., Bölte, J., & Zwitserlood, P. (2007). Describing scenes hardly seen. *Acta Psychologica*, 125(2), 129–143. <https://doi.org/10.1016/j.actpsy.2006.07.004>
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67(3), 547–619. <https://doi.org/10.1353/lan.1991.0021>
- Dowty, D. R. (1989). On the Semantic Content of the Notion of ‘Thematic Role.’ In G. Chierchia, B. H. Partee, & R. Turner (Eds.), *Properties, Types and Meaning* (Vol. 39, pp. 69–129). Springer Netherlands. [https://doi.org/10.1007/978-94-009-2723-0\\_3](https://doi.org/10.1007/978-94-009-2723-0_3)
- Ezzyat, Y., & Davachi, L. (2011). What Constitutes an Episode in Episodic Memory? *Psychological Science*, 22(2), 243–252. <https://doi.org/10.1177/0956797610393742>
- Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating Verbs, Situation Schemas, and Thematic Role Concepts. *Journal of Memory and Language*, 44(4), 516–547. <https://doi.org/10.1006/jmla.2000.2728>
- Fillmore, C. J. (1968). The Case for Case. In E. Bach, & R. T. Harms (Eds.), *Universals in Linguistic Theory* (Vol. 2, pp. 1–90). Holt, Rinehart and Winston.
- Fillmore, C. J., & Baker, C. (2009). A frames approach to semantic analysis. In B. Heine, & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis*. Oxford University Press.
- Flecken, M., Gerwien, J., Carroll, M., & von Stutterheim, C. (2015). Analyzing gaze allocation during language planning: A cross-linguistic study on dynamic events. *Language and Cognition*, 7(1), 138–166. <https://doi.org/10.1017/langcog.2014.20>

- Fuchs, T. (2017). *Ecology of the Brain: The phenomenology and biology of the embodied mind*. Oxford University Press. <https://doi.org/10.1093/med/9780199646883.001.0001>
- Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, 83(1), 49–79. [https://doi.org/10.1016/S0010-0277\(01\)00166-4](https://doi.org/10.1016/S0010-0277(01)00166-4)
- Gennari, S., & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. *Cognition*, 89(1), B27–B41. [https://doi.org/10.1016/S0010-0277\(03\)00069-6](https://doi.org/10.1016/S0010-0277(03)00069-6)
- Gerwien, J. (2011). A psycholinguistic approach to AT-structure analysis. In K. Spalek, & J. Domke (Eds.), *Sprachliche Variationen, Varietäten und Kontexte. Festschrift für Rainer Dietrich*. Stauffenburg.
- Gerwien, J., & Flecken, M. (2016). *First things first? Top-down influences on event apprehension*. 2633–2638.
- Gerwien, J., & von Stutterheim, C. (2022). 6. Describing motion events. In A. H. Jucker, & H. Hausendorf (Eds.), *Pragmatics of Space* (pp. 153–180). De Gruyter. <https://doi.org/10.1515/9783110693713-006>
- Gerwien, J., & von Stutterheim, C. (2018). Event segmentation: Cross-linguistic differences in verbal and non-verbal tasks. *Cognition*, 180, 225–237. <https://doi.org/10.1016/j.cognition.2018.07.008>
- Gerwien, J., & von Stutterheim, C. (2022). Describing motion events. In A. H. Jucker, & Heiko. Hausendorf (Eds.), *HoPs 14 Pragmatics of Space*. De Gruyter Mouton.
- Gerwien, J., von Stutterheim, C., & Rummel, J. (2022). What is the interference in “verbal interference”? *Acta Psychologica*, 230, 103774. <https://doi.org/10.1016/j.actpsy.2022.103774>
- Glanemann, R., Zwitserlood, P., Bölte, J., & Dobel, C. (2016). Rapid apprehension of the coherence of action scenes. *Psychonomic Bulletin & Review*, 23(5), 1566–1575. <https://doi.org/10.3758/s13423-016-1004-y>
- Gold, D. A., Zacks, J. M., & Flores, S. (2017). Effects of cues to event segmentation on subsequent memory. *Cognitive Research: Principles and Implications*, 2(1), 1. <https://doi.org/10.1186/s41235-016-0043-2>
- Goldin-Meadow, S. (2003). *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language*. Psychology Press.
- Goldman, A. I. (1970). *Theory of human action*. Princeton University Press.
- Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11(4), 274–279. <https://doi.org/10.1111/1467-9280.00255>
- Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, 142(3), 880–905. <https://doi.org/10.1037/a0030045>
- Hindy, N. C., Altmann, G. T. M., Kalenik, E., & Thompson-Schill, S. L. (2012). The Effect of Object State-Changes on Event Processing: Do Objects Compete

- with Themselves? *Journal of Neuroscience*, 32(17), 5795–5803. <https://doi.org/10.1523/JNEUROSCI.6294-11.2012>
- Hindy, N. C., Solomon, S. H., Altmann, G. T. M., & Thompson-Schill, S. L. (2015). A Cortical Network for the Encoding of Object Change. *Cerebral Cortex*, 25(4), 884–894. <https://doi.org/10.1093/cercor/bht275>
- Horchak, O. V., & Garrido, M. V. (2021). Dropping bowling balls on tomatoes: Representations of object state-changes during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(5), 838–857. <https://doi.org/10.1037/xlm0000980>
- Jeunehomme, O., & D'Argembeau, A. (2020). Event segmentation and the temporal compression of experience in episodic memory. *Psychological Research*, 84(2), 481–490. <https://doi.org/10.1007/s00426-018-1047-y>
- Johnson, C. R., Schwarzer-Petruck, M., Baker, C. F., Ellsworth, M., Ruppenhofer, J., & Fillmore, C. J. (2016). *FrameNet: Theory and practice*. International Computer Science Institute.
- Johnson-Laird, J. P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.
- Kang, X., Eerland, A., Joergensen, G. H., Zwaan, R. A., & Altmann, G. T. M. (2020). The influence of state change on object representations in language comprehension. *Memory & Cognition*, 48(3), 390–399. <https://doi.org/10.3758/s13421-019-00977-7>
- Kang, X., Joergensen, G. H., & Altmann, G. T. M. (2020). The activation of object-state representations during online language comprehension. *Acta Psychologica*, 210, 103162. <https://doi.org/10.1016/j.actpsy.2020.103162>
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394. <https://doi.org/10.1037/0033-295X.85.5.363>
- Klein, W. (1999). Wie sich das deutsche Perfekt zusammensetzt. *Zeitschrift für Literaturwissenschaft und Linguistik*, 29(1), 52–85. <https://doi.org/10.1007/BF03379170>
- Klein, W. (2010). On times and arguments. *Linguistics*, 48(6). <https://doi.org/10.1015/ling.2010.040>
- Kuperberg, G. R. (2021). Tea With Milk? A Hierarchical Generative Framework of Sequential Event Comprehension. *Topics in Cognitive Science*, 13(1), 256–298. <https://doi.org/10.1111/tops.12518>
- Lambert, M., von Stutterheim, C., Carroll, M., & Gerwien, J. (2022). Under the surface: A survey of principles of language use in advanced L2 speakers. *Language, Interaction and Acquisition*, 13(1), 1–28. <https://doi.org/10.1075/lia.21014.lam>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.
- Löbner, S. (2021). Cascades. Goldman's Level-Generation, Multilevel Categorization of Action, and Multilevel Verb Semantics. In S. Löbner, T. Gamerschlag,

- T. Kalenscher, M. Schrenk, & H. Zeevat (Eds.), *Concepts, Frames and Cascades in Semantics, Cognition and Ontology* (Vol. 7, pp. 263–307). Springer International Publishing. [https://doi.org/10.1007/978-3-030-50200-3\\_13](https://doi.org/10.1007/978-3-030-50200-3_13)
- Lupyan, G. (2012). Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis. *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00054>
- McRae, K., Hare, M., Elman, J. L., & Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, 33(7), 1174–1184. <https://doi.org/10.3758/BF03193221>
- Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and Perception*: Harvard University Press. <https://doi.org/10.4159/harvard.9780674421288>
- Newton, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28(1), 28–38. <https://doi.org/10.1037/h0035584>
- Paivio, A. (2014). *Mind and its evolution: A dual coding theoretical approach*. Psychology Press.
- Papafragou, A. (2015). The Representation of Events in Language and Cognition. In E. Margolis, & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 327–346). MIT Press.
- Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, 108(1), 155–184. <https://doi.org/10.1016/j.cognition.2008.02.007>
- Prystauka, Y. (2018). Comprehending Events on the Fly: Inhibition and Selection during Sentence Processing. *Master's Theses*. 1206.
- Radvansky, G. A. (2012). Across the Event Horizon. *Current Directions in Psychological Science*, 21(4), 269–272. <https://doi.org/10.1177/0963721412451274>
- Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? *Psychonomic Bulletin & Review*, 26(6), 1850–1869. <https://doi.org/10.3758/s13423-019-01634-5>
- Sauppe, S., & Flecken, M. (2021). Speaking for seeing: Sentence structure guides visual event apprehension. *Cognition*, 206, 104516. <https://doi.org/10.1016/j.cognition.2020.104516>
- Solomon, S. H., Hindy, N. C., Altmann, G. T. M., & Thompson-Schill, S. L. (2015). Competition between Mutually Exclusive Object States in Event Comprehension. *Journal of Cognitive Neuroscience*, 27(12), 2324–2338. [https://doi.org/10.1162/jocn\\_a\\_00866](https://doi.org/10.1162/jocn_a_00866)
- Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human Brain Activity Time-Locked to Narrative Event Boundaries. *Psychological Science*, 18(5), 449–455. <https://doi.org/10.1111/j.1467-9280.2007.01920.x>
- Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, 138(2), 236–257. <https://doi.org/10.1037/a0015631>

- Taylor, H., & Vickers, P. (2017). Conceptual fragmentation and the rise of eliminativism. *European Journal for Philosophy of Science*, 7(1), 17–40. <https://doi.org/10.1007/s13194-016-0136-2>
- Trueswell, J. C., & Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*, 63(1), 64–82. <https://doi.org/10.1016/j.jml.2010.02.006>
- Ünal, E., Ji, Y., & Papafragou, A. (2021). From Event Representation to Linguistic Meaning. *Topics in Cognitive Science*, 13(1), 224–242. <https://doi.org/10.1111/tops.12475>
- Van Dijk, T. A. (1995). On macrostructures, mental models, and other inventions: A brief personal history of the Kintsch-van Dijk theory. In *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 383–410).
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.
- von Stutterheim, C., Gerwien, J., Bouhaous, A., Carroll, M., & Lambert, M. (2020). What makes up a reportable event in a language? Motion events as an important test domain in linguistic typology. *Linguistics*, 58(6), 1659–1700. <https://doi.org/10.1515/ling-2020-0212>
- von Stutterheim, C., & Gerwien, J. (2023). Die Bedeutung sprachspezifischer Ereignisschemata für die Argumentstruktur. Ein Vergleich zwischen dem Ausdruck von Bewegungsereignissen im Deutschen und im Französischen. In J. Hartmann, & A. Wöllstein (Eds.), *Propositionale Argumente im Sprachvergleich | Propositional Arguments in Cross-Linguistic Research. Theorie und Empirie | Theoretical and Empirical Issues (Studien zur deutschen Sprache 84)*
- Wilson, F., Papafragou, A., Bungler, A., & Trueswell, J. C. (2011). Rapid extraction of event participants in caused motion events. *Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 33, No. 33)*.
- Wilson, V. A. D., Zuberbühler, K., & Bickel, B. (2022). The evolutionary origins of syntax: Event cognition in nonhuman primates. *Science Advances*, 8(25), eabn8464. <https://doi.org/10.1126/sciadv.abn8464>
- Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(3), 253–265. <https://doi.org/10.1002/WCS.104>
- Zacks, J. M. (2020). Event Perception and Memory. *Annual Review of Psychology*, 71(1), 165–191. <https://doi.org/10.1146/annurev-psych-010419-051101>
- Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, 4(6), Article 6. <https://doi.org/10.1038/88486>
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. In *Psychological Bulletin* (Vol. 133, Issue 2, pp. 273–293). NIH Public Access. <https://doi.org/10.1037/0033-2909.133.2.273>

- Zacks, J. M., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006). Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychology and Aging, 21*(3), 466–482. <https://doi.org/10.1037/0882-7974.21.3.466>
- Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*(1), 3–21. <https://doi.org/10.1037/0033-2909.127.1.3>
- Zuberbühler, K., & Bickel, B. (2022). Transition to language: From agent perception to event representation. *WIREs Cognitive Science, 13*(6). <https://doi.org/10.1002/wcs.1594>





---

Monique Flecken

# **The Cognitive Perspective: Language and Cognition in Event Construal**

---

How speakers of different languages talk about things happening in the world around them has been a productive area of research for many decades. Studying this from a cognitive perspective became a thriving research tradition following the work of (amongst others) Leonard Talmy (e.g., Talmy, 2000) and Dan Slobin (e.g., Slobin, 1985), who focused specifically on events involving motion, i.e., humans, animals or vehicles moving from one place to another. Motion events represent a particularly interesting test case for the study of event construal in language, given crosslinguistic differences in how speakers describe the path of motion (and other information, such as the manner of motion) in such events. Following from Talmy's 'cognitive semantics', researchers in linguistics investigated the lexical means and grammatical structures used across languages to describe motion, under the assumption that language systems reflect the conceptual structures that exist in a speaker's mind. Christiane von Stutterheim and colleagues at Heidelberg University took this as a starting point in proposing that we use these conceptual structures to make sense of the world by default. As such, they represent readily available 'scaffolds' for guiding our interpretation of activity happening in the world around us. In other words, language can be considered a window into cognition. The work by Christiane von Stutterheim, Mary Carroll, and colleagues at Heidelberg University takes a broad crosslinguistic perspective, gathering empirical data on event construal in a wide variety of languages, to investigate the extent to which crosslinguistic differences apparent in event description reflect different patterns of conceptualization. Data from speakers of different languages thus allows us to investigate the extent to which there exists diversity in linguistic encoding patterns and cognitive processes.

Von Stutterheim and colleagues' work follows a psycholinguistic framework, closely connected to the early work of Willem Levelt and colleagues (e.g., Levelt, 1989), shedding light on the processing stages that speakers go through when engaged in the act of speaking. Von Stutterheim and colleagues

took a particular interest in the cognitive processes of speakers who are planning to describe events, requiring the production of entire sentences, always with a focus on potential crosslinguistic differences during any of the processing stages involved in this task. The Heidelberg team specialized in studying the cognitive side of language production, in particular the exact (temporal) relationship between processing and the eventual speech produced, thus, the real-time link between cognitive processes during language planning and language output. In addition, cognitive processing is studied not only during language production, but also during tasks that require processing not necessarily in the realm of overt language use, with the aim of shedding light on the extent to which our speaking habits may be reflected in thinking habits, even when not using language in the moment.

The Heidelberg team advanced the research on language differences in event construal by actually studying conceptualization processes *online*, in a systematic, large-scale crosslinguistic approach. The innovation of the team's approach concerned the use of methodologies and experimental paradigms that allowed tapping into cognitive processing in real-time, in innovative and clever ways, for example, using eye-tracking methodology. The rationale behind this method lies in the fact that people's eye movements towards specific elements of a visual stimulus displaying an event reflect different stages in the conceptualization and linguistic planning process over time, starting with the initial apprehension of a scene, to the retrieval of linguistic forms from the mental lexicon in case of an event description task. Von Stutterheim et al. (2012a) is one of the first studies reporting analyses of eye movements towards elements of dynamic scenes, short video clips, especially designed for the purpose of capturing language and cognitive processing, and crosslinguistic differences therein, in motion event conceptualization. Video clips are more appropriate than hitherto used still images to study event conceptualization, as they actually depict change and dynamic activity unfolding over time, and through space, the core element distinguishing events from states, for example. These and other sets of video clips designed and shot in Heidelberg have been used extensively in follow-up research, also by other labs in the world. To date, eye-tracking studies of language production in the context of events – that is, in the realm of the production of entire sentences rather than words in isolation – are still scarce. Further, von Stutterheim and colleagues developed innovative experimental paradigms to study crosslinguistic differences in the cognitive processing of events, not necessarily in a speaking or listening mode (e.g., Flecken et al., 2014; Gerwien & von Stutterheim, 2018).

Also on the linguistic side, the Heidelberg team's focus was unique, in studying grammatical categories relevant to the expression of the temporal and spatial dimensions of an event (in particular, tense and aspect).

Grammaticalized linguistic categories are particularly important in event construal as speakers must obligatorily attend to and encode the concepts and event dimensions they denote – leading to enhanced cognitive saliency of these concepts during language production, and potentially beyond, in event processing in general. The crosslinguistic work also extended beyond looking at the construal of events in isolation to the description of entire narratives (e.g., sequences of activities unfolding over time, with a particular narrative structure, e.g., von Stutterheim & Lambert, 2005) and expository texts (von Stutterheim et al., 2012b), as well as to studying route descriptions and task instructions (von Stutterheim, 1997).

Von Stutterheim's work is particularly influential in the research areas of second language acquisition and bilingualism (see von Stutterheim & Carroll, 2011). Taking into account crosslinguistic differences for complex linguistic tasks, such as describing events or narratives in a second language, is very important, as L2 speakers often need to engage in such tasks (see von Stutterheim et al., 2013). Further, the cognitive perspective taken in this research, and the accompanying measures of cognitive processing such as the analysis of eye movements, is considered innovative and fruitful for shedding light on well-known issues such as transfer / crosslinguistic influence that can be found even in highly advanced and experienced L2 users (see overview in Lambert et al., 2022). A particularly valuable aspect of von Stutterheim and colleagues' L2 research is the focus on the 'un-usual suspects', speakers with language combinations not frequently studied in psycholinguistic research of bilingualism, such as Arabic and German (e.g., Gerwien & von Stutterheim, 2022). The research has implications also for second language teaching and didactics, as can also be seen in the contributions to this volume.

In more recent years, work in event processing in the Heidelberg lab has been extended to other areas, such as the study of clinical populations, e.g., ageing individuals with cognitive decline (early Alzheimer's; Kokje et al., 2021), patients diagnosed with schizophrenia or anorexia nervosa, in collaboration with colleagues from Heidelberg University Hospital and beyond. The interdisciplinary potential and relevance of the von Stutterheim approach for many disciplines within the cognitive sciences can be seen through the setup of numerous collaborations, involving clinicians, neuroscientists, computer scientists, artificial intelligence experts, psychologists and more.

In sum, key contributions of the work by Christiane von Stutterheim and her Heidelberg colleagues lie in investigating language and cognition, focusing on how language interacts with other cognitive processes (memory, attention, etc.), in a methodologically advanced way. In her research, various aspects of this relationship are explored, such as how language structures our thoughts, influences memory, shapes our perception of the world around us, and the accompanying constraints for second language acquisition. The

broad crosslinguistic approach taken allows for a comparison of event representations in speakers of different languages, providing valuable insights into the universal and language-specific aspects of event cognition. The present anthology illustrates the broad spectrum of research domains that Christiane von Stutterheim inspired with her work over the last 40 years.

## References

- Flecken, M., von Stutterheim, C., & Carroll, M. (2014). Grammatical aspect influences motion event perception: evidence from a cross-linguistic, non-verbal recognition task. *Language and Cognition*, 6(1), 45–78.
- Gerwien, J., & von Stutterheim, C. (2018). Event segmentation: cross-linguistic differences in verbal and non-verbal tasks. *Cognition* 180, 225–237.
- Gerwien, J., & von Stutterheim, C. (2022). Conceptual blending across ontological domains – references to time and space in motion events by Tunisian Arabic speakers of L2 German. *Frontiers in Communication: Language Science*.
- Kokje, E., Gerwien, J., & von Stutterheim, C. (2021). Macro-event recognition in healthy ageing, Alzheimer's disease, and mild cognitive impairment. *The Journal of Neuropsychology*.
- Lambert, M., von Stutterheim, C., Carroll, M., & Gerwien, J. (2022). Under the surface: A survey on principles of language use in advanced L2 speakers. *Language, Interaction and Acquisition*, 13.
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. MIT Press.
- Slobin, D. I. (Ed.). (1985). *The Crosslinguistic Study of Language Acquisition (Vol. 1: The Data; Vol. 2: Theoretical Issues)*. Lawrence Erlbaum Associates.
- von Stutterheim, C. (1997). *Einige Prinzipien des Textaufbaus. Empirische Untersuchungen zur Produktion mündlicher Texte*. (Reihe Germanistische Linguistik 184). Niemeyer.
- von Stutterheim, C., & Lambert, M. (2005). Crosslinguistic analysis of temporal perspective in text production. In H. Hendricks (Ed.), *The structure of learner varieties* (pp. 1–19). De Gruyter.
- von Stutterheim, C., & Carroll, M. (2011). The concept-oriented approach to second language acquisition. In P. Robinson (Ed.) *Routledge Encyclopedia of SLA* (pp. 110–113). Routledge.
- von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtova, B. (2012). How grammaticized concepts shape event conceptualization in the early phases of language production. Insights from linguistic analysis, eye tracking data and memory performance. *Linguistics* 50(4), 833–869.
- von Stutterheim, C., Bouhaous, A., Carroll, M., & Sahonenko, N. (2012). Language specificity and macrostructural organisation in expository texts. *Linguistics* 50(2), 341–371.

- von Stutterheim, C., Flecken, M., & Carroll, M. (2013). Principles of information organization in L2 use: Complex patterns of conceptual transfer. *International Review of Applied Linguistics in Language Teaching*, 51(2), 229–242.
- Talmy, L. (2000). *Toward a Cognitive Semantics: Volume 1. Concept Structuring Systems*. MIT Press.



---

# Autorenverzeichnis

---

**Cecilia Andorno**

Università di Torino  
ceciliamaria.andorno@unito.it

**Sandra Benazzo**

Université de Paris 8 & CNRS  
sandra.benazzo@gmail.com

**Mary Carroll**

Universität Heidelberg  
camar12@t-online.de

**Christine Dimroth**

Universität Münster  
christine.dimroth@uni-muenster.de

**Zoriana Falinska**

Universität Heidelberg  
falinska@stud.uni-heidelberg.de

**Monique Flecken**

Universiteit van Amsterdam  
m.e.p.flecken@uva.nl

**Johannes Gerwien**

Universität Heidelberg  
gerwien@idf.uni-heidelberg.de

**Michael Herweg**

Universität Heidelberg  
herweg@cl.uni-heidelberg.de

**Hana Klages**

Universität Heidelberg  
klages@idf.uni-heidelberg.de

**Wolfgang Klein**

Max Planck Institute for  
Psycholinguistics  
Wolfgang.Klein@mpi.nl

**Rolf Koeppel**

Universität Heidelberg  
rolf.koeppel@t-online.de

**Monique Lambert**

Université de Paris 8 & CNRS  
mon.lambert@wanadoo.fr

**Ines Marberg**

Universität Heidelberg  
marberg@idf.uni-heidelberg.de

**Silvia Natale**

Universität Bern  
silvia.natale@unibe.ch

**Kristian Nicolaisen**

Universität Heidelberg  
nicolaisen@idf.uni-heidelberg.de

Christiane von Steutterheim spannt in ihrer wissenschaftlichen Arbeit den Bogen von der Sprachtypologie über die Satz- und Textproduktion bis zu Fragen des Zweit- und Fremdsprachenerwerbs. Ihr besonderer Beitrag für die Forschung liegt in der Betonung der Linguistik als wichtige Teildisziplin der Kognitionswissenschaften. Die Festschrift versucht mit ihren Beiträgen die Breite von Christiane von Steutterheims Forschungstätigkeit widerzuspiegeln.



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

ISBN 978-3-96822-163-2



9 783968 221632