

Anke Holler, Thomas Weskott

# Implizite Verbkausalität im Korpus? – Eine Fallstudie

**Abstract** Experimental psycholinguists have studied the so-called implicit causality bias for more than forty years and have mostly attributed it to an effect of argument structure of a certain class of interpersonal verbs on subsequent anaphor resolution; most accounts attribute this effect to differences in salience between the arguments of the respective verbs. This article reports a corpus study on passive sentences for two classes of implicit causality verbs and puts the salience hypothesis to test. By tracing the implicit causality bias in corpora and taking into consideration a wider variety of contexts than usually employed in experiments, we want to scrutinize the ecological validity of the experimental results. From a more general point of view, the aim of the article is to exemplify how results from different methodological approaches, i.e. experiments and corpus search, can be brought to bear on our understanding of a grammatical phenomenon.

**Keywords** Implicit causality bias, psych verbs, ecological validity, passives

## 1 Einleitung

Die psycholinguistische Forschung widmet sich seit mehr als vier Jahrzehnten den beobachtbaren Effekten der sogenannten impliziten Verbkausalität. Dieses Konzept erfasst eine inhärent lexikalische Eigenschaft von transitiven Verben, die eines der beiden Argumente des Verbs (üblicherweise das AGENS) als Verursacher des ausgedrückten Sachverhalts ausweist. Damit einher geht eine erhöhte Salienz dieses Arguments, wie zahlreiche experimentelle Studien gezeigt haben. Üblicherweise wird dabei mithilfe von Satzvervollständigungsverfahren der Form [Argument<sub>1</sub>, Verb Argument<sub>2</sub>, *weil* Pronomen ...] ermittelt, welches Argument präferiert pronominal wiederaufgenommen wird. Die Präferenz bezüglich des anaphorischen Bezugs wird auch als *implicit causality bias* (kurz: IC-Bias) bezeichnet und in der Regel auf Asymmetrien in der Argumentstruktur der beteiligten Verben zurückgeführt. Diese Klasse wird in der psycholinguistischen

Literatur üblicherweise als interpersonale Verben oder IC-Verben bezeichnet. Das Hauptaugenmerk liegt aber genau genommen auf einer Teilmenge dieser Verben, die mit der linguistischen Klasse der psychischen Verben nahezu deckungsgleich ist. Obwohl dem IC-Bias eine prominente Rolle in der psycholinguistischen Literatur zukommt, ist die Frage bisher nicht thematisiert worden, ob und inwieweit die beobachteten Effekte des IC-Bias ökologisch valide sind, d. h. außerhalb eines sorgfältig kontrollierten Experiments Bestand haben. Da die Datenerhebung zudem bisher nur in Experimentalstudien mit einer sehr eingeschränkten Menge von Stimuli erfolgt ist, wurden mögliche grammatische Einflüsse auf den IC-Bias kaum berücksichtigt. Nennenswerte Ausnahmen stellen die Arbeiten von Corrigan (1988) zur Belebtheit oder von Brown & Fish (1983) zur Definitheit dar.

Korpora haben den Vorzug, dass sie gebrauchsbasiert sind und daher für eine sprachliche Zielgröße u. a. Auskunft über ihr Vorkommen, die Häufigkeit ihres Vorkommens und/oder die sprachliche Umgebung ihres Vorkommens geben können. Unter der Annahme, dass Korpusdaten somit einen direkteren Zugriff auf das multivariate Zusammenspiel grammatischer Eigenschaften erlauben als experimentelle Befunde, wollen wir untersuchen, ob sich der IC-Bias in Korpora abbilden lässt. Dabei wollen wir anhand einer Fallstudie zur impliziten Verbkausalität in passivierten Sätzen zeigen, dass Korpusdaten zur Überprüfung der ökologischen Validität von Experimentaldaten herangezogen werden können. Mit anderen Worten: Wir möchten der Frage nachgehen, inwieweit Korpusdaten geeignet sind, aufzuklären, ob sich bestimmte sprachliche Entitäten im Gebrauch genauso verhalten wie im wohlkontrollierten Experiment. Wir nehmen an, dass nur dann, wenn die Antwort auf diese Frage positiv ausfällt, es gerechtfertigt ist, die Ergebnisse einer experimentellen linguistischen Untersuchung als ökologisch valide einzustufen. Unser Vorschlag ist also, Korpusfrequenzdaten zu nutzen, um für experimentelle Settings in der Linguistik den Grad ihrer Approximation an wirkliche sprachliche Gegebenheiten zu bestimmen. Der Aufsatz schließt damit an die grundsätzliche Diskussion darüber an, wie Ergebnisse, die durch verschiedene methodische Zugänge zustande kommen, zusammengenommen für die Analyse eines grammatischen Phänomens geltend gemacht werden können. In unserem konkreten Fall heißt dies, die bestimmten psychischen Verben eigene Salienzmarkierung ihrer Argumente, die experimentell im IC-Bias ihren Ausdruck findet, auch durch korpusbasierte Gebrauchsdaten zu belegen.

Mit unserer Studie verfolgen wir ausdrücklich nicht das Ziel, die linguistische Theoriebildung fortzuentwickeln und beispielsweise zu grundlegenden Fragen hinsichtlich der Motivation für Passivierung oder zur sprachtheoretischen Behandlung von psychischen Verben beizutragen, sondern es geht uns vorrangig um ein methodologisches Problem. Indem wir exemplarisch aufzeigen, wie mittels eines im Korpus leicht zugänglichen linguistischen Ausdruckstyps (i. e. der

Passivierung) ein psycholinguistischer Befund (i. e. der IC-Bias) validiert werden kann, wollen wir einen Beitrag zur Beantwortung der generellen Frage leisten, in welcher Weise Korpusdaten und Experimentaldaten einander sinnvoll komplementieren können.

Der Aufsatz gliedert sich wie folgt: Nach einer ausführlichen Beschreibung des IC-Bias, seiner psycholinguistischen Fundierung und der sich daraus ergebenden Forschungsfrage bezüglich der ökologischen Validität der Experimentaldaten im nachfolgenden Abschnitt 2 werden wir in Abschnitt 3 die Korpusstudie vorstellen, die wir durchgeführt haben, um die in der Psycholinguistik zur Erklärung des IC-Bias gängige Salienzhypothese anhand von Korpusfrequenzdaten zu überprüfen. Im abschließenden Abschnitt 4 werden wir die Ergebnisse dieser Studie problematisieren und zur Diskussion über die ökologische Validität der psycholinguistischen Befunde in Beziehung setzen.

## 2 Fragestellung

Im Zuge der empirischen Wende in der Sprachwissenschaft haben experimentell erhobene sprachliche Daten für die linguistische Theoriebildung an Bedeutung gewonnen. Verglichen mit anderen linguistischen Datentypen zeichnen sie sich vor allem dadurch aus, dass sie aus kontrolliert durchgeführten Experimenten stammen, in denen sprachliche Phänomene hypothesengeleitet hinsichtlich einer oder mehrerer zuvor festgelegter Kriterien nach einer mehr oder minder verbindlich vorgegebenen Prozedur untersucht werden. Auf diese Weise ist im Idealfall weitestgehend sichergestellt, dass die bezüglich des jeweiligen sprachlichen Phänomens beobachteten Effekte tatsächlich auf die untersuchten Faktoren bezogen werden können und nicht etwa Resultat anderer, zuvor nicht berücksichtigter Einflussgrößen sind. Gleichzeitig stellt sich aber auch die Frage, inwieweit die so gewonnenen Daten überhaupt die realen Gegebenheiten repräsentieren. Eine weitergehende Frage ist dann, ob experimentelle Bedingungen die Gegebenheiten des natürlichen Sprachgebrauchs approximieren können und wollen. Demgegenüber liefern Korpora Evidenz, die nicht aus einer kontrollierten experimentellen Situation stammt und damit nicht aus einer Vorauswahl der zu betrachtenden Eigenschaften resultiert; vielmehr kann in Korpusstudien eine Zufallsauswahl an Textbelegen getroffen werden, ohne im Vorhinein die Anzahl der Eigenschaften festlegen oder auch nur kennen zu müssen. Insofern ist naheliegend zu fragen, ob beide Datentypen, Experimentaldaten und Korpusdaten, sinnvoll miteinander kombiniert werden können, sodass insgesamt in Bezug auf ein sprachliches Phänomen ein vollständigeres und verlässlicheres Bild entsteht. Dieses Zusammenspiel von experimentell beobachteten Effekten einerseits und korpusbasiert erhobenen Befunden andererseits wollen wir anhand des IC-Bias

einer Subklasse der Psychverben eingehender betrachten. In einem ersten Schritt werden wir dazu nachfolgend einige methodologische Aspekte, insbesondere hinsichtlich der ökologischen Validität und des IC-Bias, diskutieren und danach die Salienzhypothese für den IC-Bias, die als Ausgangspunkt für die in Abschnitt 3 dargestellte Korpusstudie dient, einführen sowie die von uns verwendete Operationalisierung mittels der Passivkonstruktion motivieren.

## 2.1 Experiment vs. Korpus

Neben explorierenden, hypothesengenerierenden Experimenten sind es vor allem hypothesentestende Experimente, denen in der Linguistik und Psycholinguistik eine gewichtige Rolle zukommt. Bei diesem Experimenttyp soll anhand einer Stichprobe von Sprechern und von (morphologischen, syntaktischen, semantischen) Instanzen eines Ausdruckstyps (den *Items*) eine Hypothese geprüft werden oder, genauer gesagt, es soll die zugehörige Nullhypothese anhand der Daten der Stichprobe verworfen und damit die empirische Hypothese angenommen werden. Der entscheidende Schritt ist dabei die Inferenz vom Nicht-Zutreffen der Nullhypothese (zum Beispiel: Es besteht hinsichtlich der Mittelwerte einer abhängigen Variablen  $v$  kein Unterschied zwischen den beiden Ausprägungen eines Faktors  $A$ ,  $a_1$  und  $a_2$ , d. h.  $\bar{x}_{v(a_1)} = \bar{x}_{v(a_2)}$ .) in der Stichprobe auf das Zutreffen der empirischen Hypothese (zum Beispiel: Es besteht ein Unterschied zwischen den Mittelwerten,  $\bar{\mu}_{v(a_1)} \neq \bar{\mu}_{v(a_2)}$  in der Population.), bei dem die Irrtumswahrscheinlichkeit üblicherweise auf 5 % festgelegt ist. Die Kombination von systematischer Manipulation eines Faktors bei gleichzeitiger Minimierung des Einflusses von Störvariablen durch Kontrolle über die experimentelle Situation mit inferenzstatistischen Verfahren stellt sicher, dass der Unterschied zwischen den erhobenen Mittelwerten der abhängigen Variablen mit einer Irrtumswahrscheinlichkeit von  $p < .05$  tatsächlich auf den Faktor zurückzuführen ist, genauer gesagt, dass die Wahrscheinlichkeit, dass das Unterschiedsmuster beobachtbar ist, wenn in der Population die Nullhypothese gilt, kleiner 5 % ist. Bei den meisten inferenzstatistischen Verfahren handelt es sich dabei um sogenannte *parametrische* Verfahren, d. h. es werden anhand von Stichprobenparametern (wie beispielsweise dem Mittelwertunterschied und dessen Streuung beim  $t$ -Test) bestimmte in der Population geltende Parameter (der in der Population geltende Mittelwertunterschied sowie dessen Streuung) geschätzt.

Ein Nachteil dieser Kombination von kontrolliertem Experiment und hypothesenprüfender Inferenzstatistik ist, dass der Anzahl der Faktoren, die in einem Experiment manipuliert werden können, durch die Anforderungen der statistischen Verfahren und deren Auswirkung auf Itemanzahl und damit letztlich auf die Belastbarkeit der Probanden recht enge Grenzen gesetzt sind: Um

eine verlässliche Schätzung der Populationsparameter für eine experimentelle Bedingung zu erhalten (wie beispielsweise eines Ratingmittelwertes und seiner Streuung über Probanden bzw. Items hinweg), braucht man für diese Bedingung mindestens 6 Beobachtungen (d.h. 6 Probanden/Items pro Bedingung). Für ein Experiment, das zwei zweistufige Faktoren miteinander kreuzt, ergibt sich also nach dieser Faustregel schon die Notwendigkeit, 24 Probanden und 24 Items zu testen; zusammen mit einer weiteren „Goldenen Regel“ des Experimentierens, und zwar, dass das Verhältnis von Filleritems (d.h. Ablenkern von der experimentellen Fragestellung) zu experimentellen Items 2:1 sein sollte, ergäben sich daraus schon  $24 + 48 = 72$  Items, die von 24 Probanden bearbeitet werden müssen. Jeder weitere zweistufige Faktor, der hinzugenommen wird, verdoppelt diese Anforderungen; spätestens bei einem Experiment, das die gegenseitigen Abhängigkeiten von vier zweistufigen Faktoren testet (also, in der Redeweise der Experimentalpsychologie, bei einem  $2 \times 2 \times 2 \times 2$ -Design), erschöpfen sich – im wahren Sinne des Wortes – die Möglichkeiten der experimentellen Überprüfung in *einem* Experiment. Die Anzahl der zu bearbeitenden Items übersteigt dann (bei ca. 300 Items) bereits die Menge, die Probanden üblicherweise zugemutet werden kann, ohne dass die Datenqualität durch Ermüdungserscheinungen leidet.

Demgegenüber haben Korpusstudien den Vorteil, dass sich satzbasiert große Mengen von Daten hinsichtlich einer potenziell beliebigen Anzahl von Faktoren annotieren lassen und es damit durchaus möglich ist, um beim Beispiel zu bleiben, gleich große Substichproben für die 16 Zellen eines  $2 \times 2 \times 2 \times 2$ -Designs zu erhalten. Anhand einer solchen Stichprobe für ein gegebenes Phänomen ließe sich dann, ganz analog zum inferenzstatistischen Verfahren beim Experimentieren, von einem Befundmuster für die Stichprobe (die annotierten Korpusbelege) auf die Population (das Korpus als Ganzes oder gar das Genre, das das Korpus repräsentiert) schlussfolgern. Allerdings wäre hier eine Anzahl von  $n=6$  Beobachtungen pro Ausprägungen eines Annotationsmerkmals schon deshalb nicht hinreichend, weil in Korpusbelegen – mangels Kontrolle über die Faktoren – außer den annotierten Eigenschaften immer noch weitere Eigenschaften zwischen den Items variieren; um eine systematische Konfundierung mit diesen Eigenschaften zu vermeiden, muss die Stichprobengröße entsprechend angepasst werden, was im Falle seltener Phänomene (d.h. lexikalisch und/oder syntaktisch restringierter Vorkommen der Form) problematisch sein kann.

Zusammenfassend lässt sich konstatieren, dass experimentelle Methoden aufgrund der systematischen Kontrolle von Faktoren Generalisierbarkeit um den Preis der Natürlichkeit des jeweiligen experimentellen Settings erkaufen, während korpusanalytische Methoden weitestgehend natürliche Daten liefern, allerdings um den Preis der systematischen Kontrolle der Einflussgrößen.

## 2.2 Ökologische Validität

Jede für ein Experiment gewählte Verfahrensweise hat letztlich ihre Grenzen im Einfluss der jeweiligen Versuchspersonen und der situativen Umgebung. Auch durch eine möglichst wirklichkeitsnahe Gestaltung eines experimentellen Settings können diese Einflussgrößen nur bis zu einem bestimmten Grad minimiert werden. Insofern stellt sich für jedes experimentell gewonnene Ergebnis die Frage nach seiner Generalisierbarkeit, die letztlich auch davon abhängt, inwieweit die Experimentalanordnung mit den natürlichen Gegebenheiten übereinstimmt. Dieser Zusammenhang ist in der Psychologie mit dem Konzept der ökologischen Repräsentativität bzw. ökologischen Validität erfasst worden, die auch als „die empirische Gültigkeit einer psychologischen Aussage für das Alltagsgeschehen“ (Dorsch 2013) beschrieben wird. Es geht also um das Verhältnis zwischen der konstruierten und in diesem Sinne artifiziellen Datenerhebung im Labor und den Gegebenheiten in einer natürlichen Lebensumgebung. Bezogen auf sprachliche Daten ist also zu fragen, worin die für die Gültigkeit der Daten relevanten natürlichen Lebensumstände bestehen und wie ihre Bedingungen bestimmt werden können. Auf den ersten Blick ist es naheliegend, den alltäglichen Sprachgebrauch als das für die ökologische Validität relevante „Biotop“ (Pawlik 1976) anzusehen. Doch um das Ausmaß der Entsprechung und damit die Vorhersageleistung der kontrollierten Datenerhebungen in der jeweiligen Kritisierungssituation ermitteln zu können, bedarf es einer Methode, die es ermöglicht, die „unverstellten“ natürlichen Sprachdaten als Vergleichsgrundlage zu archivieren. Dies kann durch die Erstellung und Aufbereitung von großen Korpora geleistet werden. (Eine Analogie besteht hier möglicherweise zu der Unterscheidung zwischen nicht-intervenierenden/beobachtenden vs. experimentellen Verfahren in der Psychologie.) Da Korpusdaten, die üblicherweise auch als Produktionsdaten beurteilt werden, aus sprachlichem Handeln in einer konkreten Situation resultieren, schlagen wir vor, Korpora auszuwerten, um die empirische Repräsentativität von experimentell gewonnenen Sprachdaten zu überprüfen. Allerdings können Korpora den sprachlichen Alltag nur approximieren, da Korpusdaten im Zuge der Auswahl und Aufbereitung bereits kategorisiert und interpretiert werden. Hinzu kommt, dass in Korpora die Textsorte „Zeitungstext“ dominiert und die dort enthaltenen Äußerungen in den meisten Fällen editiert sind. Ein vollkommen unverstellter, natürlicher Zugang zu Sprachdaten, wie es das Desiderat der ökologischen Validität erfordert, ist damit auch durch Korpora nicht vollständig gegeben. (Um die Analogie zur Psychologie aufzugreifen: Dies ist mutatis mutandis vergleichbar mit dem Beobachterparadoxon in der (Sozial-) Psychologie.) Unter der Annahme, dass Korpusdaten einen direkteren Zugriff auf grammatische Gegebenheiten erlauben als experimentelle Befunde, scheinen Korpora aber gut geeignet, um die Übertragbarkeit von Laborergebnissen

auf tatsächliche sprachliche Äußerungssituationen festzustellen und damit die ökologische Validität beobachteter Effekte zu bestimmen. Inwieweit dieser Zusammenhang hergestellt werden kann, wollen wir in einer Fallstudie ergründen. Anhand des IC-Bias interpersonalen Verben, eines in der Psycholinguistik fest etablierten Befundes, werden wir untersuchen, inwieweit er sich in Korpora abbilden lässt und welchen Rückschluss dies auf seine ökologische Validität zulässt.

### 2.3 IC-Bias der psychischen Verben

Unter dem Etikett *implizite Kausalität* wird in der experimentellen Psychologie und der Psycholinguistik das Phänomen verhandelt, dass die Information, die der Zuschreibung von Ursache und Wirkung in Ereignissen zugrunde liegt (die kausale Attribution zu Partizipanten eines Ereignisses), in den lexikalischen Einträgen von Verben implizit enkodiert ist. Für agentiv-kausative *accomplishment-* oder *achievement-*Verben wie beispielsweise *schlagen* lässt sich diese Zuschreibung ohne Weiteres aus der syntaktischen Funktion der Argumente ablesen: Der Referent des Subjekts des SCHLAGEN-Ereignisses ist – qua AGENS – ursächlich für die Wirkung, welche im Affiziertsein des Referenten des direkten Objekts durch das SCHLAGEN-Ereignis (als Bestandteil des Resultatzustandes) besteht. Weniger klar und nicht aus der syntaktischen Funktion der Argumente ableitbar ist die Zuschreibung dieser kausalen Rollen<sup>1</sup> für psychische Verben. Den paradigmatischen Fall bildet das Paar *fürchten* vs. *ängstigen*:

- (1) a. Siggie ängstigt Erwin.  
b. Erwin fürchtet Siggie.

In einer Situation, in der Siggie Erwin ängstigt, ist es wahrscheinlich gleichzeitig der Fall, dass Erwin Siggie fürchtet. Hieraus ergibt sich schon, dass die kausale Rolle des Verursachers nicht am Subjekt festgemacht werden kann; vielmehr scheint die Rolle des Verursachers in (1.a) beim Referenten des Subjekts, in (1.b) aber bei dem des direkten Objekts verortbar zu sein (also beide Male bei Siggie); die Wirkung, also der psychische Zustand des Fürchtens bzw. Geängstigtheits, ist in (1.a) im Objekt-, in (1.b) aber im Subjektreferenten zu finden (also beide Male bei Erwin). Es liegt nahe, die kausalen Rollen an die jeweiligen thematischen Rollen zu knüpfen: der die Furcht/Angst erfahrende

1 Die Redeweise von Kausalität ist hier selbstverständlich nicht im Sinne naturgesetzlicher Determiniertheit der Wirkung durch die Ursache zu verstehen.

oder erlebende Erwin ist in beiden Fällen eher Teil der Wirkung, während der Furcht/Angst einflößende Siggì eher Teil der Ursache ist. Dies mag Garvey & Caramazza (1974), die diese Beobachtung erstmals experimentell absicherten, dazu bewogen haben, von *causal valence* zu sprechen. In neuerer Terminologie ist es bei psychischen Verben die Rolle des STIMULUS, der man kausale Wirkmächtigkeit, und die des EXPERIENCERS, der man die (psychische) Affiziertheit durch das Verbereignis zusprechen würde (siehe Postal, 1971; Belletti & Rizzi, 1988 und die darauffolgende Literatur). Wir schließen uns dieser Nomenklatur an und werden, wie das in der psycholinguistischen Literatur zum Thema üblich ist, psychische Verben wie *ängstigen*, *nerven*, *überraschen* im Folgenden als STIMULUS-EXPERIENCER-Verben (kurz: SE-Verben) bezeichnen, und psychische Verben wie *fürchten*, *hassen*, *bemerkten* als EXPERIENCER-STIMULUS-Verben (kurz: ES-Verben).<sup>2</sup>

Garvey & Caramazza (1974) kommt das Verdienst zu, auf die Asymmetrie hinsichtlich der kausalen Valenz aufmerksam gemacht zu haben; ihr Squib in *Linguistic Inquiry* kann als Anfangspunkt eines regelrechten Industriezweigs der Psycholinguistik und der experimentellen Sprach- und Sozialpsychologie gesehen werden. Die Asymmetrie der kausalen Valenz bei psychischen Verben ist in den letzten etwa vierzig Jahren für unzählige Sprachen und Sprechergruppen (Sprachlerner, Erwachsene) nachgewiesen worden, und zwar mithilfe verschiedenster experimenteller Paradigmen sowohl psycholinguistischer wie auch kognitions- und sozialpsychologischer Provenienz: vom einfachen *Rating* der kausalen Wirkmächtigkeit für die beiden Referenten über Attributions- und Satzvervollständigungsaufgaben bis zu komplexen Online-Erhebungsmethoden wie *Eyetracking* beim Lesen und im *Visual World*-Paradigma (siehe Pickering & Majid 2007 und Hartshorne 2013 für einen Überblick). Dabei erwies sich vor allem ein Effekt als äußerst robust, der aus (psycho-)linguistischer Sicht interessant ist: Präsentiert man Probanden hinsichtlich eines Pronomens wie *er* ambige Satzfragmente der Form in (2) und bittet sie, diese zu vervollständigen, so zeigt sich, dass in den Satzvervollständigungen das ambige Pronomen *er* koreferent mit dem STIMULUS-Argument (*Siggì*) ist.

- (2) a. Siggì ängstigte Erwin, weil er ...  
 ... schon wieder die Wildschweinmaske trug.  
 (*er* = Siggì<sub>STIMULUS</sub>)

2 Es sei der Leserin überlassen, diese durch die in der syntaktischen Literatur gängige Terminologie zu ersetzen, derzufolge *ängstigen* ein OBJECT-EXPERIENCER-Verb, *fürchten* aber ein SUBJECT-EXPERIENCER-Verb ist.

- b. Erwin fürchtete Siggi, weil er ...  
 ... schon wieder die Wildschweinmaske trug.  
 (*er* = Siggi<sub>STIMULUS</sub>)

Diese Präferenz zugunsten des STIMULUS-Arguments bei der Anaphernresolution ist als ein empirisches Korrelat des IC-Bias in die Literatur eingegangen. Naturgemäß war es eher der psycholinguistische als der sozialpsychologische Forschungszweig, der sich diesem Befund gewidmet hat, und zwar nicht zuletzt deshalb, weil Fälle wie (2.b) offensichtlich den gängigen Heuristiken zuwiderlaufen, die im Sprachverstehen wie in der Sprachproduktion für die Anaphernresolution angesetzt werden: Subjektpräferenz und First-Mention-Strategie, denen zufolge das Antezedens von *er* in (2.b) *Erwin* sein müsste (qua Subjekt bzw. erst-erwähntem Referenten).

Der IC-Bias zugunsten des STIMULUS-Arguments bei der Anaphernresolution beläuft sich den Metaanalysen von Ferstl, Garnham & Manoulidou (2011) und Hartshorne (2013) zufolge auf ca. 85 % (mit Schwankungen in Abhängigkeit vom betrachteten Verb bzw. Verbpaar). In der psycholinguistischen Literatur besteht weitestgehend Einigkeit darüber, dass es die Asymmetrie auf der Ebene der Argumentstruktur ist, die für den IC-Bias verantwortlich ist; siehe dazu v. a. Hartshorne (2013), der überzeugend darlegt, dass es sich beim IC-Bias um eine grammatisch getriebene und nicht, wie in der Sozialpsychologie (zum Beispiel von Rudolph & Försterling 1997) angenommen, eine weltwissensbedingte Präferenz handelt.

Worin besteht nun diese Asymmetrie? Es ist prima facie nicht unbedingt einleuchtend, dass ein anaphorischer Prozess zwischen den thematischen Rollen STIMULUS und EXPERIENCER unterscheidet – üblicherweise wird angenommen, dass anaphorische Prozesse wie die Resolution eines (ambigen) Pronomens von Faktoren wie (linearer) Distanz, syntaktischer Bindung (im Falle von intrasententialen Vorkommen) und Diskursrelation (im Falle von intersententialen Anaphern) abhängen (siehe dazu z. B. Garnham & Cowles 2008; Kehler & Rohde 2013). Dies scheint zumindest für das Beispiel in (2) nicht der Fall zu sein: Lineare Distanz, syntaktische Gegebenheiten und Diskursrelation sind über die beiden Varianten hinweg konstant. Legt man die – in der Literatur mehr oder minder akzeptierte – Annahme zugrunde, dass die Anaphernresolution beim Sprachverstehen im Falle zweier möglicher Antezedenten denjenigen auswählt, dessen Repräsentation im Arbeitsgedächtnis der Hörerin/Leserin das höhere Aktivationsniveau hat, ergibt sich als Erklärung für den IC-Bias die folgende Salienzhypothese<sup>3</sup> (kurz: SH):

3 Wir sind uns im Klaren darüber, mit Salienz eine (kognitive) Größe in die Hypothese einzuführen, deren linguistische Reflexe und Rolle in der Verarbeitung notorisch

(SH) In einem Kontext der Form [NP<sub>1</sub> Verb NP<sub>2</sub>, *weil* PersPron] wird das ambige Personalpronomen präferiert zugunsten der NP aufgelöst, die vom psychischen Verb die salientere Argumentrolle zugewiesen bekommt. Für psychische Verben ist dies unabhängig von der syntaktischen Funktion die STIMULUS-Rolle.

Explizit vertreten wird die Salienzhypothese im Zusammenhang mit dem IC-Bias von Kasof & Lee (1993), wie das folgende Zitat illustriert:<sup>4</sup>

Because people attribute greater causality to more salient stimuli than to less salient stimuli, people reading sentences implying different levels of salience for subjects and objects should attribute the interpersonal events unequally between subjects and objects. To the degree that a sentence evokes a mental representation in which the subject is more salient than the object, readers should attribute the interpersonal event more to the subject than to the object. To the degree that a sentence evokes a mental representation in which the object is more salient than the subject, readers should attribute the interpersonal event more to the object than to the subject. (ebd., p.878).

Diese Hypothese, die – mehr oder minder explizit – von den meisten Autoren, die zum IC-Bias arbeiten, geteilt wird, macht also eine Salienzasymmetrie auf der Ebene der Argumentrollen verantwortlich für die Abhängigkeit der Resolutionspräferenz für ambige Pronomina in *weil*-Fortsetzungen vom Typ des psychischen Verbs. Mit anderen Worten: Präsentiert man Probanden unvollständige Satzfolgen, bestehend aus einem Satz mit zwei genusidentischen NPen, die als Argumente eines psychischen Verbs fungieren, und einem mit *weil* und ambigem Personalpronomen eingeleiteten Satzfragment, so vervollständigen sie diese Satzfolgen präferent mit dem STIMULUS-Argument des psychischen Verbs des ersten Satzes, weil das STIMULUS-Argument salienter und damit das geeignetere Antezedens für das Personalpronomen ist.

unklar sind; dies geschieht zugegebenermaßen in Ermangelung eines weniger vagen Konzepts mit spezifischeren empirischen Korrelaten. Der Begriff der Prominenz scheint uns in diesem Zusammenhang nicht weniger problembehaftet. Siehe hierzu auch Vogel (2015).

4 *Stimulus* wird hier von den Autoren im Sinne von ‚in einem Experiment dargebotener Reiz‘, nicht im Sinne einer Theta-Rolle verwendet.

## 2.4 Zur Operationalisierung der Salienzhypothese im Korpus

Wie in Abschnitt 2.3 ausführlich dargestellt, wird in der Psycholinguistik als gängige Erklärung für den IC-Bias ein Salienzunterschied zwischen den Argumenten der in Rede stehenden Verben angesetzt. Experimentell wird das Salienzgefälle zwischen den Argumenten zumeist durch die Auswertung anaphorischer Bezüge ermittelt, wobei in der Regel angenommen wird, dass Probanden (beispielsweise in Satzvervollständigungsaufgaben) ein im nachfolgenden Satzfragment vorhandenes Pronomen auf das jeweils salientere Argument im Vorgängersatz beziehen. Insofern wäre es naheliegend, auch im Korpus anaphorische Bezüge auszuwerten und zu diesem Zwecke vorhandene Koreferenzannotationen zu nutzen. Abgesehen davon, dass Zweifel berechtigt sind, ob in den vorhandenen für Koreferenz annotierten Korpora überhaupt genügend Belege für die uns interessierenden psychischen Verben auffindbar wären, ist ein solches Vorgehen mit einem vergleichsweise hohen Aufwand verbunden. Die Korpora könnten nicht rein satzbezogen ausgewertet werden; vielmehr müsste für alle Fälle, in denen die untersuchten IC-Verben vorkommen, der jeweilige Kontext auf mögliche anaphorische Größen hin überprüft werden, die das Subjekt oder das Objekt des jeweiligen Verbs wiederaufnehmen. Dabei stellte sich nicht nur die Frage nach den berücksichtigten Formen (z. B. nur Pronomen oder auch komplexe NPen), sondern auch die Frage nach der Größe des Suchfensters: Wie viele Sätze können zwischen dem Zielsatz und dem Satz, der den anaphorischen Ausdruck enthält, liegen, damit der anaphorische Bezug für die Fragestellung noch relevant ist, und wie kann dieses über die Fälle hinweg normiert werden? Abgesehen von diesen Schwierigkeiten in der Umsetzung spricht gegen die Auswertung von anaphorischen Bezügen im Korpus aber vor allem, dass zur ökologischen Validierung der psycholinguistischen Salienzhypothese für den IC-Bias unabhängige Korpusevidenz nötig ist. Da das anaphorische Potenzial aber bereits in den psycholinguistischen Experimenten als abhängige Variable für die Ermittlung von Salienzunterschieden genutzt wurde, wäre dies mit der Auswertung von anaphorischen Bezügen über Koreferenzannotationen in Korpora nicht mehr ohne Weiteres gegeben. Vielmehr bedarf es eines anderen unabhängigen Maßes, um die experimentellen Ergebnisse korpusbasiert validieren zu können.

Welches Maß, das an der syntaktischen Oberfläche zugänglich ist, wäre nun geeignet, um Salienzunterschiede zwischen sprachlichen Ausdrücken im Korpus zu ermitteln? Salienz lässt sich im Korpus nur vermittelt und in Bezug auf grammatische Eigenschaften oder Relationen zwischen Ausdrücken feststellen, von denen unabhängig bekannt ist, dass sie die Salienz eines Ausdruckes bzw. genauer: seines Diskursreferenten erhöhen. Dies trifft beispielsweise auf Ersterwähnungen, Subjekte, Topiks usw. zu. Rein von der syntaktischen Oberfläche

aus gesehen sind diese funktionalen Ausdrücke aber unterschiedlich leicht ablesbar; am leichtesten ließen sich anhand der Kasusmarkierung wohl noch die Subjekte bestimmen. Um aber die Salienzhypothese zu überprüfen, bedarf es für die Korpusrecherche einer abhängigen Variable, die nicht nur einen nachvollziehbaren Schätzwert für die Salienz darstellt, sondern auch im Korpus beobachtbar, d.h. suchbar und zählbar ist. Passivkonstruktionen scheinen dieser Anforderung gerecht zu werden. Um die ökologische Validität der psycholinguistischen Salienzhypothese zu überprüfen, bedienen wir uns daher der passivierten Form von SE- und ES-Verben. Durch die Passivierung kann bei SE-Verben das STIMULUS-Argument als optionales PP-Argument auftreten, während bei ES-Verben das EXPERIENCER-Argument als optionales PP-Argument realisiert würde.<sup>5</sup> Nach Hypothese SH enthält diese optionale PP bei SE-Verben das salientere Argument, nicht jedoch bei ES-Verben. Diesen Zusammenhang nutzen wir für die Operationalisierung der Salienzhypothese im Rahmen unserer Korpusstudie. Wir nehmen an, dass die Realisierung des PP-Arguments im Passiv vom Salienzstatus des Referenten, der mit diesem Argument verbunden ist, abhängig ist, und erwarten, dass es für in diesem Sinne salientere Argumente wahrscheinlicher ist, als PP realisiert zu werden als für weniger saliente Argumente, weswegen STIMULUS-Argumente häufiger als PP-Argument im Passiv vorkommen sollten als EXPERIENCER-Argumente. Wird ein Satz mit einem SE-Verb passiviert (vgl. 3.a), sollten mehr overte PPen vorkommen als bei der Passivierung eines Satzes mit einem ES-Verb (vgl. 3.b).

- (3) a. Erwin wird (von Siggi) geängstigt.  
 b. Siggi wird (von Erwin) gefürchtet.

Das Auftreten der PP im Passiv erscheint uns zum einen wegen seiner Optiona-  
 lität und zum anderen wegen seiner problemlosen Überführbarkeit in Häufigkeiten  
 als abhängige Variable besonders gut geeignet.

5 Die Frage nach der Motivation für die Passivierung blenden wir hier aus, auch weil es abgesehen von einem Verweis auf die mit der Subjektfunktion verbundene Prominenz und dem mit der Passivierung einhergehenden Perspektivenwechsel keine einfache Antwort darauf gibt. Was immer aber die Passivierung auslösen mag, diese Gründe bleiben konstant in Bezug auf die betrachteten Sätze mit oder ohne *von*-PP. Den Terminus *von*-PP verwenden wir hier als Oberbegriff für mit *von* und anderen, idiosynkratischen Präpositionen gebildeten PPen.

### 3 Korpusstudie

#### 3.1 Design und Vorhersage

Überträgt man die aus der psycholinguistischen Literatur ableitbare Hypothese hinsichtlich des Salienzunterschiedes in Satzvervollständigungsexperimenten auf die hier interessierende Fragestellung hinsichtlich des Vorkommens der *von*-PP in Passivierungen, so lautet die empirische Vorhersage  $H_1$  (komplementär zur Nullhypothese  $H_0$ ):

( $H_1$ ): STIMULUS-Argumente werden in Passivstrukturen häufiger overt in einer *von*-Phrase realisiert als EXPERIENCER-Argumente, da erstere salienter sind als letztere.

Die zugehörige Nullhypothese lautet entsprechend, dass sich kein Unterschied in der Häufigkeit der overten Realisierung zwischen den beiden Argumenttypen findet. Die wichtigste unabhängige Variable war der zweistufige Faktor VERBTYP (SE- vs. ES-Verb); abhängige Variable war die relative Häufigkeit des Auftretens des im Passivsatz nicht als Subjekt realisierten Arguments in der *von*-Phrase. Dieses  $1 \times 2$ -Design wurde zusätzlich um den Zufallsfaktor VERBPAAR erweitert, um eine Generalisierung über verschiedene SE/ES-Paare zu ermöglichen; Details hierzu werden im folgenden Abschnitt dargelegt.

#### 3.2 Methode

Im Sinne der Zielvorgabe der Ermittlung der ökologischen Validität experimenteller Befunde zu IC-Verben wurden in einem ersten Schritt Verbpaare von STIMULUS-EXPERIENCER- und EXPERIENCER-STIMULUS-Verben identifiziert, für die experimentell erhobene Satzvervollständigungsdaten vorliegen und die über mehrere Experimente hinweg einen reliablen Unterschied gezeigt haben, der im Sinne der Salienzhypothese interpretierbar ist.<sup>6</sup> Das Kriterium der Paarbildung war dabei, dass eine Einsetzung konkreter Verben und ihrer Argumente in das folgende Schema eine plausible Beschreibung eines (kausalen) Zusammenhanges liefert:

6 Wir haben uns dabei an den Bias-Daten aus eigenen Satzvervollständigungsexperimenten sowie an den Bias-Daten von Ferstl, Garnham & Manoulidou (2011) zum Englischen orientiert. Dass der IC-Bias crosslinguistisch robust ist, zeigen die Arbeiten von Bott & Solstad (2014) und Hartshorne & Snedeker (2010).

(S): Wenn  $x$   $y$  SE-verbt, dann ist es möglich, dass  $y$   $x$  ES-verbt.

Das Paradebeispiel für eine Einsetzung in dieses Schema mit plausiblen Resultat ist das Paar *fürchten/ängstigen*: Wenn Peter Hans ängstigt, dann ist es möglich – und vielleicht sogar kausal erklärbar –, dass Hans Peter fürchtet. Dass sich dieses Schema nicht über alle Verbpaare gleich gut durchhalten ließ, zeigen Einsetzungsbeispiele wie *Wenn Peter Hans begeistert, dann ist es möglich, dass Hans Peter bewundert.* oder noch etwas weniger plausibel: *Wenn Peter Hans beunruhigt, dann ist es möglich, dass Hans Peter verdächtigt.* Entscheidend für diesen Test war letztlich, dass die Einsetzung des Verbpaares in das Schema nicht zu völliger Unplausibilität führt; das war für keines der ermittelten Verbpaare der Fall. Es ergaben sich die folgenden Verbpaare: *fürchten/ängstigen*; *bewundern/begeistern*; *bedauern/erschüttern*; *ertragen/stören*; *verachten/enttäuschen*; *hassen/beleidigen*; *bemerkern/überraschen* und *verdächtigen/beunruhigen*. Die Aufgabe bei der Erstellung des Korpus war, für diese Verbpaare ein Korpus von je 200 Belegen von Passivsätzen (100 pro Verb) zu erstellen, was die Verwendung eines großen Korpus notwendig machte. Darüber hinaus sollten die Belege idealerweise innerhalb der Paare und über die Paare hinweg hinsichtlich ihrer Genrezugehörigkeit kontrolliert (d. h. *gematcht*) sein.

*Extraktion der Stichprobe.* Da DEREKO für die fraglichen Verbpaare im Passiv nicht hinreichend große Stichproben lieferte, haben wir das DWDS-Kernkorpus (vgl. Geyken 2007) genutzt, das gleichzeitig über ein Genrefeature verfügt. Wir verwendeten das über die URL [eins.dwds.de](http://eins.dwds.de) zugängliche „alte“ Graphic User Interface des DWDS, das die Belege nach Textsorten sortiert ausgibt, da dies eine kontrollierte Extraktion von Belegen in Abhängigkeit vom Genre ermöglichte. Die Suchanfrage lautete `<“ge-V-t” @werden>` – es wurden also alle Fälle des Partizips II mit adjazenter Form des Hilfsverbs extrahiert; darunter fanden sich einige Dubletten, Futur- und Konjunktivformen, die von Hand aussortiert werden mussten. Der Suche nach adjazenten Wörtern lag keine theoretische, sondern eine rein praktische Überlegung zugrunde, da die Suche nach nicht-adjazenten Wortformen (`<“ge-V-t” && “werden”>`) zwar deutlich mehr Belege, aber eben auch mehr falschpositive Formen (z. B. Futur- und Konjunktivformen) zeitigte. Extrahiert wurden durch diese lemmabasierte Suchanfrage auch orthografisch abweichende Formen wie z. B. *geängstiget wardt*. Tabelle 1 gibt einen Überblick über die Anzahl von Belegen.

Unser Bestreben, innerhalb der Verbpaare eine möglichst balancierte Verteilung hinsichtlich des Faktors GENRE zu haben, war leider nicht bei allen Verbpaaren erfolgreich. Für Verbpaarlinge mit wenigen Belegen (d. h. etwa 120 Fälle, davon im Mittel ca. 20 Falschpositive) waren wir auf die Genrezusammensetzung dieses Verbs festgelegt; der andere Paarling musste dann, soweit möglich, der Genrezusammensetzung des ersten Paarlings angepasst werden. Abbildung 1

Tabelle 1: Anzahl extrahierter Fälle von Passivformen nach VERBTYP.

SE-Verben	N		ES-Verben	N
ängstigen	109		fürchten	307
begeistern	145		bewundern	872
erschüttern	1.758		bedauern	175
stören	3.056		ertragen	386
enttäuschen	1.592		verachten	403
beleidigen	626		hassen	111
überraschen	2.192		bemerkn	973
beunruhigen	118		verdächtigen	411
TOTAL	9.596			3.638

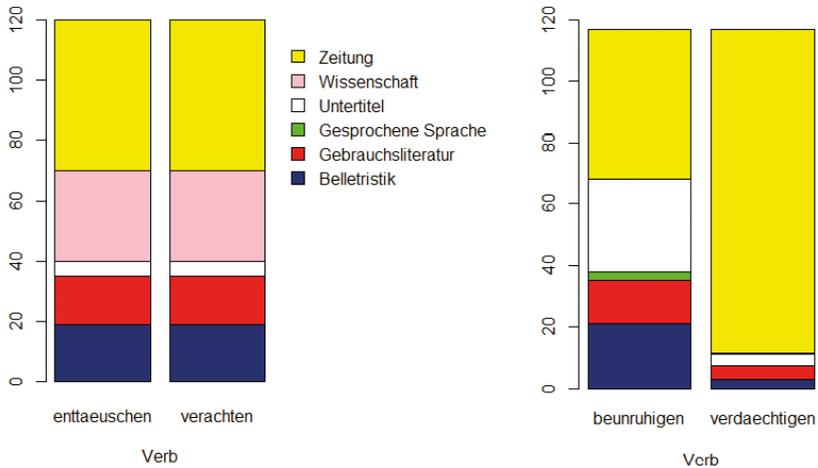


Abbildung 1: Genrezusammensetzung für die Verbpaare *enttäuschen/verachten* und *beunruhigen/verdächtigen*.

zeigt zwei extreme Fälle, die durch dieses Vorgehen entstanden sind: Das Verbpaar *enttäuschen/verachten* ließ sich hinsichtlich GENRE vollständig balancieren; die maximale Abweichung zwischen zwei Paarlingen mussten wir beim Verbpaar *beunruhigen/verdächtigen* in Kauf nehmen.

In einigen wenigen Fällen mussten nach Bereinigung der extrahierten 120 Belege noch Belege „nachgezogen“ werden, da zu viele Dubletten und Falschpositive enthalten waren. Nach diesem Schritt lagen für jeden der 16 Verbpaarlinge 100 Passivformen vor.

*Datenaufbereitung.* Die extrahierten und bereinigten 1.600 Fälle wurden mithilfe des Statistik-Software-Pakets R (R Core Team 2014, Version 3.1.1) so aufbereitet, dass jeder der Belege, die in ihrer Länge und Komplexität massiv variierten, in einer Tabelle einheitlich wie folgt repräsentiert war: Die Partizipialform bildete eine eigene Spalte, der diese Form umgebende Text (soweit vorhanden) die linke und rechte Spalte daneben. Vor diese Spalten wurde eine Spalte mit einer durchlaufenden Zählvariable, eine Spalte mit dem Verb und eine Spalte mit der Ausprägung des Faktors VERBTYP gesetzt. Hinter der Spalte mit dem Text, der der Partizipialform folgte, wurden die zu annotierenden Merkmale jeweils spaltenweise angeordnet.

*Annotation.* Die Annotation erfolgte in MS Excel anhand folgender Merkmale:<sup>7</sup>

- VERBTYP: SE- vs. ES-Verb (automatisch erzeugt)
- PASSIV: Passivform vs. keine Passivform ([0,1]-kodiert)
- VON.PP: Realisierung eines der beiden Argumente in einer *von*-Phrase (oder einer äquivalenten PP; [0,1]-kodiert, wie auch alle folgenden Merkmale)
- IDIO.PP: idiosynkratische, vom Verb selegierte Präpositionen (z.B. *geängstigt werden durch*; [0,1]-kodiert)
- IDIO.PP.FORM: Wortform der idiosynkratischen PP
- S.ANIM: Belebtheitsstatus des Subjekts
- PP.OBJ.ANIM: Belebtheitsstatus des PP-Objekts
- S.DEF: Definitheitsstatus des Subjekts
- PP.OBJ.DEF: Definitheitsstatus des PP-Objekts
- POSITION: relative Position von Subjekt und PP-Objekt; die unmarkierte Abfolge wurde als „0“ kodiert.

Die Annotation des Merkmals, das für unsere Fragestellung zentral ist, VON.PP, erwies sich als vollständig unproblematisch. Hier zwei Beispiele für Kodierung mit „1“ und „0“; im Folgenden wird die Verbform unterstrichen und die *von*-PP **in Fettdruck** wiedergegeben:

- (4) Auch in diesem Jahr dürfte der Kreisverkehr, der selbst **von versierten Automobilisten** gefürchtet wird, seine Spitzenstellung halten – wenn nicht bald etwas geschieht.<sup>8</sup>

7 Hier möchten wir für die Unterstützung bei der Annotation der Daten Miriam Feix, Cheryl Hodgkinson, Heinke Jank und Markus Paluch danken.

8 DWDS, Subkorpus *Berliner Zeitung*; Textklasse: Zeitung::Lokales; Berliner Zeitung vom 01.07.1997.

- (5) Zum 13. Mal erzielte Polster in einer Bundesligapartie zwei Treffer – gefürchtet wird er unter dem Künstlernamen Toni Doppelpack.<sup>9</sup>

Auch die Annotation des Merkmals IDIO.PP war, von einigen wenigen Zweifelsfällen abgesehen, unproblematisch; diese Zweifelsfälle – handelt es sich bei einer idiosynkratischen PP tatsächlich um das Argument, das dem Subjekt im Aktivsatz entspricht? – wurden im Kreis der Annotatoren diskutiert und geklärt. Hier ein Beispiel für IDIO.PP==1 und IDIO.PP.FORM==mit:

- (6) Ist der Darwinismus des Marktes, **mit dem** wir täglich geängstigt werden, nur das Spiel einer Elite?<sup>10</sup>

Größere Schwierigkeiten bereitete – wenig überraschend – die Annotation der Belebtheitsmerkmale S.ANIM und PP.ANIM. In Zweifelsfällen wie NPen, die Gremien, Institutionen und sonstige Kollektive von Individuen bezeichnen (*der Aufsichtsrat, das Parlament, die Studierendenschaft*) optierten wir im Zweifel für den Merkmalswert „belebt“, da es die individuellen psychischen Zustände und Absichten der in die Summenindividuen eingehenden Elemente sind (die Aufsichtsräte, die Parlamentarier, die Studierenden), die Träger der EXPERIENCER- bzw. STIMULUS-Eigenschaften sind. Ein Beispiel für einen unzweifelhaften Fall von S.ANIM==0 gibt (7):

- (7) In der Regel bezieht sich die fixe Idee auf unerreichte Zwecke, auf Güter, die gehofft, auf Uebel, die gefürchtet werden.<sup>11</sup>

Der Definitheitsstatus der beiden Argumente bereitete keine größeren Annotationsprobleme. Allerdings traten Fälle von koordinierten NPen auf, die gemischten Belebtheits- oder Definitheitsstatus aufwiesen; hier ein Beispiel für letzteren:

- (8) So haben wir gesehen, daß der große Komet des Jahres 1456 Entsetzen über ganz Europa verbreitete, das ohnehin schon **durch**

9 DWDS, Subkorpus *Berliner Zeitung*; Textklasse: Zeitung::Sport; Berliner Zeitung vom 26.08.1996.

10 DWDS, Subkorpus *Berliner Zeitung*; Textklasse: Zeitung::Magazin; Berliner Zeitung vom 06.12.1997.

11 DWDS, Subkorpus *Deutsches Textarchiv*; Textklasse: Wissenschaft::Medizin; Reil, Johann Christian (1803), Rhapsodien über die Anwendung der psychischen Curmethode auf Geisteszerrüttungen. Halle.

**eine verheerende Pest und durch die Verwüstungen, welche die Türken um sich verbreiteten, geängstigt wurde.**<sup>12</sup>

In Fällen wie (8) wurde das Merkmal PP.DEF mit „9“ kodiert, was in der statistischen Analyse als fehlender Wert interpretiert wurde.

Linguistisch interessanter waren Problemfälle, die sich aus der Split-Stimulus-Konstruktion (siehe Engelberg 2015) ergaben; (9) gibt ein Beispiel:

- (9) Die Ausschreitungen von damals seien „bis heute für Rostock ein Brandmal“, sagte Gauck, dessen Rede kurzzeitig **von Zwischenrufen wie „Heuchler“ durch Linksautonome gestört wurde.**

Anhand von (9) lassen sich einige der Annotationsprobleme nochmals illustrieren: Das Subjekt des passivierten Relativsatzes ist das EXPERIENCER-Argument *dessen* [also Gaucks] *Rede*; dies wurde folglich als S.ANIM==0 und S.DEF==1 annotiert. Das STIMULUS-Argument allerdings ist zweigeteilt. Gaucks Rede wird *von Zwischenrufen durch* Linksautonome gestört, d. h. die VON-PP ist unbelebt, die IDIO.PP aber belebt.

Da die Wortstellung von Subjekt und *von*-PP im Satz höchst selten markiert war, verursachte auch diese Annotation keine Probleme.

Zur Erinnerung: Wir sagen, der Salienzhypothese folgend, vorher, dass sich die abhängige Variable – relative Auftretenshäufigkeit der VON-PP bzw. IDIO-PP – in Abhängigkeit vom Faktor VERBTYP unterschiedlich verhalten sollte. Die Fälle ohne *overt* PP sollten im Falle passivierter ES-Verben häufiger sein als im Falle von SE-Verben, weil bei letzteren das (per Hypothese salientere) STIMULUS-Argument der Kandidat für die Realisierung der PP ist. Und umgekehrt, und aus demselben Grund, sollten die Fälle von *overt* realisierter PP häufiger bei SE- als bei ES-Verben auftreten. Statistisch sagen wir also eine (disordinale) Interaktion der Faktoren VERBTYP und REALISIERUNG DER PP vorher.

12 DWDS, Subkorpus *Deutsches Textarchiv*; Textklasse: Gebrauchsliteratur::Populärwissenschaft::Wissenschaft::Physik; Littrow, Joseph Johann von (1836): *Die Wunder des Himmels, oder gemeinfaßliche Darstellung des Weltsystems*. Bd. 3. Stuttgart.

### 3.3 Ergebnisse

Die annotierten 1.600 Fälle verteilen sich wie folgt auf die sich aus der Kreuzung der beiden Faktoren ergebenden vier Zellen:

Tabelle 2: Absolute Häufigkeiten der Realisierung der PP in Abhängigkeit vom Faktor VERBTYP.

	Verbtyp	
	ES	SE
ohne overte PP	583	408
mit overter PP	217	392

Abbildung 2 illustriert dieses Befundmuster grafisch.

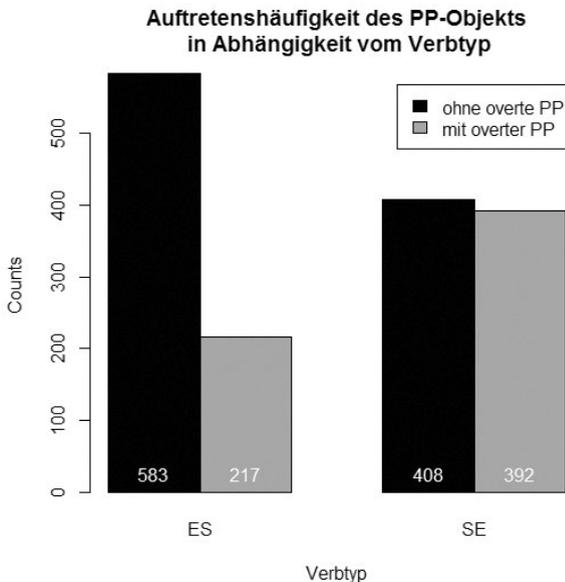


Abbildung 2: Absolute Häufigkeiten von Passivsätzen mit overter und ohne overte PP in Abhängigkeit vom Verbtyp.

Um das sich numerisch andeutende Interaktionsmuster – 466 Fälle Unterschied zwischen overter vs. non-overter PP für ES-Verben und nur 16 Fälle Unterschied für SE-Verben – inferenzstatistisch zu erhärten, wurde eine *mixed-model* logistische Regression auf der [0,1]-kodierten abhängigen Variablen *Realisierung der PP* gerechnet; diese ergab sich aus dem Aufaddieren der mit „1“ kodierten Fälle von

realisierten *von*-PPen und idiosynkratischen PPen.<sup>13</sup> Die Berechnung der logistischen Regression wurde in R mit dem Paket *lme4* (Version 3.1, Bates et al. 2015) und dem Befehl für generalisierte lineare Modelle, *glmer*, durchgeführt. Die acht Verbpaare behandelten wir dabei als Zufallsfaktor (*random factor*) und VERBTYP als festen (*fixed factor*).

Dieses Verfahren schätzt den Einfluss des Prädiktors VERBTYP auf die logit-transformierte abhängige Variable *Realisierung der PP*. Das heißt, es gibt uns eine Antwort auf die Frage, wie viel der im Datensatz vorhandenen Varianz dem manipulierten Faktor zuzuschreiben ist und wie viel davon reine Fehlervarianz („Rauschen“) ist. Die Fehlervarianz kann bei dem von uns gewählten Verfahren nochmals durch die Berücksichtigung des Messwiederholungsfaktors VERBPAAR unterteilt werden: Ist der Einfluss des Faktors VERBTYP signifikant (d.h. lässt er eine Generalisierung auf die „Population“ von Sätzen zu, aus der wir unsere Stichprobe gezogen haben), so sollte sich dieser Einfluss innerhalb der Verbpaare und über diese hinweg zeigen. Die logit-Transformation der abhängigen Variablen wird dabei durchgeführt, um eine  $\alpha$ -Fehler-Inflation zu vermeiden, die bei der Berechnung des Modells auf untransformierten absoluten Häufigkeiten droht (siehe Agresti 2002). Ein signifikantes Ergebnis für eine logistische Regression besagt in unserem Fall also, kurz gesagt, dass wir die Nullhypothese, dass der untersuchte Faktor VERBTYP keinen Einfluss auf die Auftretenswahrscheinlichkeit der PP hat, verwerfen und die Alternativhypothese (VERBTYP hat einen Einfluss) annehmen dürfen. Der *p*-Wert gibt uns dabei die Wahrscheinlichkeit an, mit der wir ein der Alternativhypothese entsprechendes Datenmuster (oder ein extremeres) finden können, obwohl in der Population das der Nullhypothese entsprechende Datenmuster gilt. Wir folgen den geltenden Standards und nehmen ein  $\alpha$ -Fehler-Niveau von .05 an; der ermittelte *p*-Wert sollte also unter diesem kritischen Wert liegen.<sup>14</sup>

Die logistische Regression ergab einen signifikanten Einfluss des Faktors VERBTYP auf die Häufigkeit der Realisierung der PP ( $\beta = 1.01$ ,  $|z| = 2.76$ ,  $p = .04$ ) – der Einfluss des Faktors VERBTYP auf die Gesamtvarianz im Datensatz ist also reliabel nachweisbar: Das Nicht-Subjekt-Argument eines passivierten psychischen Verbs tritt reliabel häufiger auf, wenn es sich um ein STIMULUS-Argument als wenn es sich um ein EXPERIENCER-Argument handelt.

13 Ein Pearson- $\chi^2$ -Test ist bei diesem Design wegen der Abhängigkeit der Faktorstufen nicht möglich.

14 Die Modellgleichung der *mixed model* logistischen Regression lautete wie folgt: `glmer(PPA~verdtype+(1+verdtype|item), data=d.clean, family=binomial, REML=FALSE)`. Dieses Modell wurde mit dem Nullmodell (d.h. `PPA~1+(1+verdtype|item)` ...) verglichen; vgl. zu diesem Vorgehen Barr et al. (2013). Weitere Details der statistischen Analyse werden hier aus Platzgründen ausgespart; Rohdaten und R-Skripte stellen wir auf Anfrage gern zur Verfügung.

*Post-hoc-Analyse.* Nimmt man allerdings weitere Prädiktoren wie beispielsweise die Definitheit des Subjekts und dessen Belebtheit ins Modell auf, über deren Einfluss wir keine Hypothese formuliert haben und die damit als *post-hoc*-Faktoren anzusehen sind, so sinkt der Beitrag von VERBTYP zur Varianzaufklärung in den marginal signifikanten Bereich ( $p = .08$  bei Hinzunahme von S.ANIM als Faktor und  $p = .11$  bei Hinzunahme von S.ANIM und S.DEF.)<sup>15</sup>

Das bedeutet, dass man zur Vorhersage der Häufigkeit der Realisierung einer *von*-PP in einem Satz mit passiviertem psychischen Verb den Faktor VERBTYP zwar heranziehen kann, und eine relativ gute Vorhersage des Auftretens der PP innerhalb der Verbpaare und über die Verbpaare hinweg erhält. Ein guter Teil dieser Vorhersagekraft dieses Faktors speist sich aber offenbar aus anderen Faktoren, deren wirkmächtigste der Belebtheitsstatus und die Definitheit des Subjekts sind.

Dies wirft die Frage auf, wie sich die abhängige Variable aus unserer Korpusstudie zu den aus psycholinguistischen Satzvervollständigungsexperimenten stammenden IC-Bias-Werten verhält. Da beide Variable [0,1]-kodierte sind, lassen sie sich ohne Transformation auf einer Skala abtragen; Abbildung 3 stellt den direkten Vergleich der Daten aus den beiden Evidenzquellen grafisch dar. Wie man anhand der unterschiedlichen Steigungen der Geraden in Abbildung 3 sieht, ist der experimentell erhobene Bias stärker als seine Entsprechung in der Korpusstudie.

In einem weiteren Schritt kann man den Zusammenhang zwischen den Korpus- und den experimentellen Daten auf Ebene der einzelnen Verbpaare betrachten; dies soll Abbildung 4 leisten:

Wie Abbildung 4 zeigt, ist der experimentell erhobene Bias über die Verbpaare hinweg relativ stabil. Der Effekt des Faktors VERBTYP variiert auf dieser Variablen zwar in seiner Stärke (d. h. der Steigung der Geraden, die die beiden Mittelwerte verbindet), aber das Vorzeichen der Steigung ist stets positiv. Das gilt für die Korpusdaten nicht. Hier gibt es zwei Verbpaare mit negativer Steigung auf der abhängigen Variablen (*hassen/beleidigen* und *verachten/enttäuschen*) sowie ein Verbpaar nahezu ohne Effekt, d. h. einer Geraden ohne Steigung (*bewundern/begeistern*). Durch unsere *post-hoc*-Tests haben wir mindestens zwei der Faktoren (Definitheits- und Belebtheitsstatus des Subjekts) identifiziert, denen wir diese Abweichung zwischen den beiden Evidenzquellen zuschreiben können.

15 Die Hinzunahme des Belebtheits- und Definitheitsstatus des PP-Objekts führte nicht zu einer signifikanten Veränderung der Varianzaufklärung. Im Modellvergleich (Likelihood Ratio  $\chi^2$ -Test, siehe Barr et al. 2013), dem derzeit wahrscheinlich konservativsten Test für gemischte Modelle, liegt der  $p$ -Wert für den Vergleich des komplexesten Modells – mit VERBTYP, S.ANIM und S.DEF – mit dem Vergleichsmodell – mit S.ANIM und S.DEF – bei  $p = .12$ .

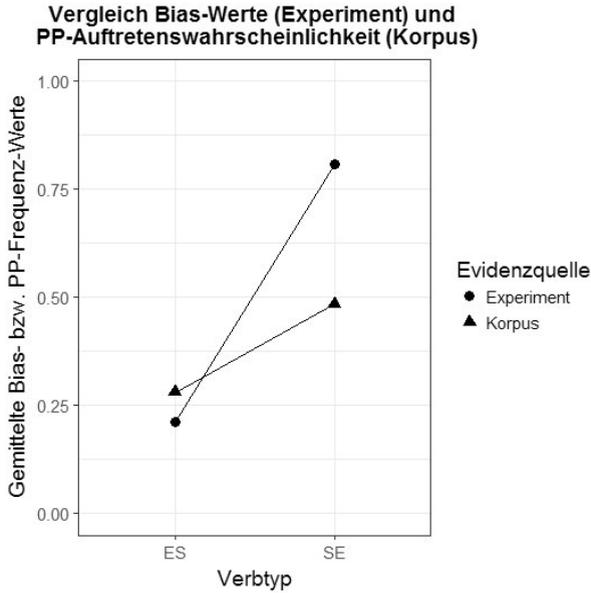


Abbildung 3: Vergleich der Mittelwerte von experimentell erhobenem Anaphernresolutions-Bias (●) und der Auftretenswahrscheinlichkeit der PP in unserem Passivkorpus (▲).

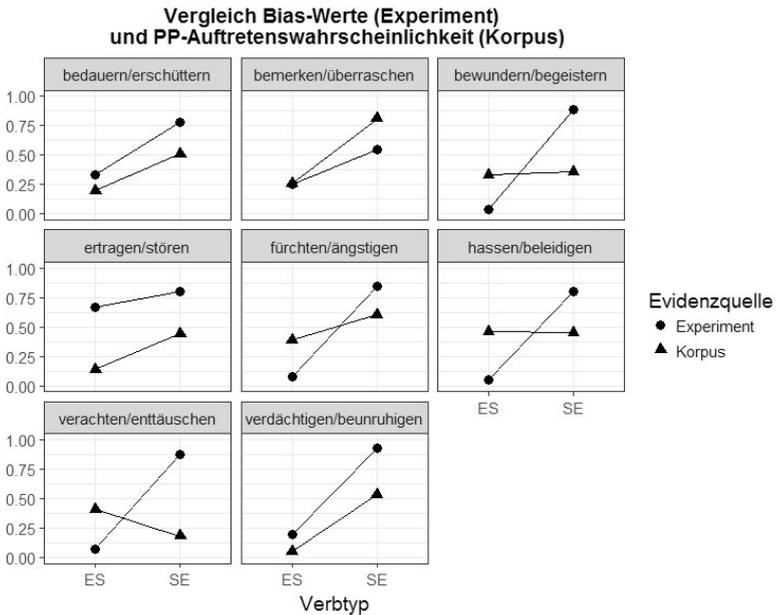


Abbildung 4: Vergleich der Mittelwerte von experimentell erhobenem Anaphernresolutions-Bias (●) und der Auftretenswahrscheinlichkeit der PP in unserem Passivkorpus (▲).

Selbstverständlich wären weitere Einflussgrößen denkbar, die die Abweichung zwischen Experiment und Korpus erklärbar machen, insbesondere informationsstrukturelle und diskursbezogene Faktoren. Diese zu erfassen hätte allerdings den ohnehin beträchtlichen Annotationsaufwand noch vergrößert, und ihre Einbeziehung in eine statistische Analyse des Auftretens der *von*-PP in Passivsätzen mit Psychverben bleibt zukünftigen Untersuchungen vorbehalten.

#### 4 Diskussion

Die Befunde der logistischen Regression für den Faktor haben gezeigt, dass der üblicherweise in psycholinguistischen Experimenten manipulierte Faktor, die Zugehörigkeit des Verbs zur SE- vs. zur ES-Klasse, die Auftretenswahrscheinlichkeit von PPs in Sätzen mit passivierten psychischen Verben verlässlich vorhersagen kann. Allerdings zeigten die *post-hoc*-Analysen, dass der Einfluss dieses Faktors nicht unabhängig zu sehen ist von Eigenschaften der Argumente, die in Satzvervollständigungsexperimenten üblicherweise *nicht* manipuliert werden: Der Definitheits- und der Belebtheitsstatus des Subjekts sind starke Prädiktoren für das Auftreten der *von*-PP – und sie sind in ihrem Potenzial zur Aufklärung von Varianz in den Korpusdaten deutlich stärker als der Faktor VERBTYP. In psycholinguistischen (Satzvervollständigungs-)Experimenten werden zumeist Eigennamen als Argumente der Verben dargeboten, d.h. belebte definite Ausdrücke. Eine Ausnahme ist in dieser Hinsicht die Studie von Corrigan (1988), die den Belebtheitsstatus der Argumente in drei Experimenten systematisch mit dem Faktor VERBTYP gekreuzt hat und signifikante Interaktionen dieser Faktoren sowohl für Kausalitätsratings als auch für Satzvervollständigungen gefunden hat. Während bei SE-Verben auch inanimaten Referenten von STIMULUS-Argumenten eine kausale Rolle im vom Verb denotierten Ereignis attribuiert wird, ist dies bei ES-Verben in weitaus geringerem Ausmaß der Fall. Und auf unbelebte STIMULUS-Referenten von SE-Verben wird in Satzvervollständigungen häufiger Bezug genommen als auf die Referenten unbelebter Stimuli von ES-Verben. Worauf Corrigans Daten hindeuten, ist eine Korrelation zwischen verbspezifischen Belebtheitsrestriktionen und den abhängigen Variablen typischer psycholinguistischer Experimente zu IC-Verben – eine Korrelation, die von den meisten nachfolgenden Studien ignoriert wurde.<sup>16</sup> Unsere Korpusdaten deuten in eine ähnliche Richtung. Während SE-Verben offenbar weniger anfällig für den Effekt der Belebtheit des Subjekts (d.h., im Passiv, des STIMULUS-Argumentes) sind, hat

16 So stellen Kasof & Lee schon 1993 fest: „Prior research on implicit causality has centered on a rather narrow range of sentence forms.“ (ebd., S. 878). Leider hat sich daran auch seitdem nicht viel geändert.

die Animateheit des Subjekts auf die ES-Verben einen deutlich stärkeren Einfluss. Die stärkere Streuung, die die verschiedenen SE-Verben in Abbildung 4 hinsichtlich der Korpusdaten aufweisen, lässt sich – tentativ und *post hoc* – über den Einfluss der verbspezifischen Animateheitsanforderung bei diesen Verben erklären: je weniger stark die Belebtheitsanforderung an das STIMULUS-Argument, desto niedriger der Score der Korpusdaten. Dies sei an einem Beispiel illustriert. Das STIMULUS-Argument eines STÖREN- oder ÜBERRASCHEN-Ereignisses ist hinsichtlich seines Belebtheitsstatus möglicherweise weniger festgelegt als ein ENTTÄUSCHEN- oder BELEIDIGEN-Ereignis. Während der STIMULUS des Gestörtwerdens ebenso gut ein Geräusch sein kann wie eine Person, die das Geräusch produziert, ist es schwerer vorstellbar, dass einem Beleidigtwerden ein unbeleibter STIMULUS zugrunde liegt. Eine systematische quantitative Analyse des Einflusses dieser Animateheitsprofile (also: animat-animat, inanimat-inanimat, animat-inanimat, inanimat-animat) der Verben ist aufgrund der Ungleichverteilung der Profile in unserem Korpus nicht möglich. Es scheint aber lohnend, solche Analysen anhand größerer Korpora zu erstellen, um diese Faktoren und ihr Zusammenspiel in balancierten Subkorpora zu untersuchen. Wir können somit festhalten, dass der Belebtheitsstatus eine geringere Rolle spielt, wenn das STIMULUS-Argument im Aktiv als Subjekt des Satzes realisiert wird.

Damit schließt die Diskussion unmittelbar an die Frage nach der ökologischen Validität der experimentell erhobenen Daten zum IC-Bias an. Diese ist offenbar eher niedrig, weil die Faktoren, die in den psycholinguistischen Experimenten im Regelfall konstant gehalten werden, de facto einen großen Einfluss auf die beobachtete Argumentasymmetrie der hier untersuchten psychischen Verben haben können. Der Preis des kontrollierten Vorgehens beim Experiment, welches notwendig zum Ausschluss möglichst vieler denkbarer Störvariablen führt, besteht also im Informationsverlust über ebenfalls vorhandene grammatische Einflussgrößen, die die Ausprägung der abhängigen Variablen potenziell modulieren. Insofern lässt sich Schütze (2006) nur beipflichten, wenn er schreibt: „Because no single kind of data is perfect, an efficacious approach to linguistic investigation is to seek converging evidence from a wide array of types of data whenever possible.“ Die hier vorgestellte Fallstudie sehen wir als weiteren Beleg für unsere Überzeugung, dass empirische Studien in der Linguistik an Aussagekraft gewinnen, wenn sie auf konvergenten Ergebnissen beruhen, die durch verschiedene Methoden erzielt wurden. Dies eröffnet zugleich die Möglichkeit, Korpora zur Validierung experimenteller Daten zu nutzen. Während experimentelle Studien kontrolliert sind und daher immer nur Ausschnitte des jeweils untersuchten Phänomens beleuchten können, repräsentieren Korpora das gesamte Spektrum der möglichen Einflussgrößen, allerdings konfundiert, d. h. in Korpusanalysen können diese Faktoren nicht kontrolliert werden. Das ermöglicht es aber, Korpora zur ökologischen Validierung und damit ggf. als Korrektiv

für experimentell erhobene Ergebnisse heranzuziehen. Darüber hinaus eröffnet diese Strategie die Möglichkeit, im Korpus identifizierte Einflussgrößen – wie in unserem Fall Animatheit und Definitheit – als Faktoren in kontrollierte Experimente aufzunehmen und dort einer systematischen, d. h. nicht-konfundierten Untersuchung ihres Effekts zuzuführen.

Zusammenfassend lässt sich konstatieren, dass die beiden betrachteten Methoden – Korpusanalyse und Experiment – gerade wegen der Komplementarität ihrer jeweiligen Stärken und Schwächen letztlich immer aufeinander bezogen und idealerweise systematisch parallelisiert durchgeführt werden sollten.

## Literaturverzeichnis

- Agresti, Alan (2002): *Categorical Data Analysis*. Hoboken, NJ: Wiley.
- Barr, Dale J./Levy, Roger/Scheepers, Christoph/Tily, Harry J. (2013): Random effects structure for confirmatory hypothesis testing: Keep it maximal. In: *Journal of Memory and Language* 68(3), S. 255–278.
- Bates, Douglas/Mächler, Martin/Bolker, Benjamin M./Walker, Steven C. (2015): Fitting Linear Mixed Effects Models Using lme4. In: *Journal of Statistical Software* 67(1), doi: 10.18637/jss.v067.i01.
- Belletti, Adriana/Rizzi, Luigi (1988): Psych Verbs and Theta Theory. *Natural Language and Linguistic Theory* 6, S. 291–352.
- Bott, Oliver/Solstad, Torgim (2014): From Verbs to Discourse: A novel account of implicit causality. In: Hemforth, Barbara/Mertins, Barbara/Fabricius-Hansen, Cathrine (Hg.): *Meaning and Understanding across Languages. Studies in theoretical psycholinguistics*. Chicago: Springer International, S. 213–251.
- Brown, Roger/Fish, Deborah (1983): The psychological causality implicit in language. In: *Cognition* 14, S. 237–273.
- Corrigan, Roberta (1988): Who dun it? The influence of actor-patient animacy and type of verb in the making of causal attribution. In: *Journal of Memory and Language* 27, S. 447–465.
- Engelberg, Stefan (2015): Gespaltene Stimulus-Argumente bei Psych-Verben. Quantitative Verteilungsdaten als Indikator für die Dynamik sprachlichen Wissens über Argumentstrukturen. In: Engelberg, Stefan/Meliss, Meike/Proost, Kristel/Winkler, Edeltraud (Hg.): *Argumentstruktur – Valenz – Konstruktionen*. Tübingen: Narr, S. 469–491.
- Ferstl, Evelyn/Garnham, Alan/Manouilidou, Christina (2011): Implicit causality bias in English: a corpus of 300 verbs. In: *Behavior Research Methods* 43(1), S. 124–135.
- Garvey, Cathrine/Caramazza, Alfonso (1974): Implicit causality in verbs. In: *Linguistic Inquiry* 5, S. 459–464.

- Garnham, Alan/Wind Cowles, Heidi (2008): Looking both ways: The JANUS model of noun phrase anaphor processing. In: Gundel, Jeanette K./Hedberg, Nancy (Hg.): *Reference: Interdisciplinary perspectives*. Oxford: Oxford University Press, S. 246–272.
- Hartshorne, Joshua K. (2013): What is implicit causality? In: *Language, Cognition and Neuroscience* 29(7), S. 804–824.
- Kasof, Joseph/Lee, Ju Young (1993): Implicit Causality as Implicit Salience. In: *Journal of Personality and Social Psychology* 65(5), S. 877–891.
- Kehler, Andrew/Rohde, Hannah (2013): A Probabilistic Reconciliation of Coherence-Driven and Centering-Driven Theories of Pronoun Interpretation. In: *Theoretical Linguistics* 39, S. 1–37.
- Pawlik, Kurt (1976): Ökologische Validität: Ein Beispiel aus der Kulturvergleichsforschung. In: Kaminski, G. (Hg.): *Umweltpsychologie. Perspektiven – Probleme – Praxis*. Stuttgart: Klett, S. 59–72.
- Pickering, Martin J./Majid, Asifa (2007): What are implicit causality and implicit consequentality? In: *Language and Cognitive Processes* 22, S. 780–788.
- Postal, Paul M. (1971): *Cross-Over Phenomena*. New York: Holt, Rinehart and Winston, S. 39ff.
- R Core Team (2014): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. URL <http://www.R-project.org/>.
- Schütze, Carston T. (2006): Data and Evidence. In: Brown, Keith (Hg.): *Encyclopedia of Language and Linguistics*. 2. Aufl., Bd. 3, Oxford: Elsevier, S. 356–363.
- Vogel, Ralf (2015): Is prominence a useful concept for the theory of syntax? *Lingue e Linguaggio*, 14(1), 97–112. doi:10.1418/80756.
- Wirtz, Markus Antonius (Hg.) (2013): *Dorsch – Lexikon der Psychologie*. 16. Aufl., Verlag Hans Huber.